Semiparametric Joint Models to Assess the Effect of Timing of the Indication for Treatment on Survival with Application to Cancer Screening

by

Wenjia Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2021

Doctoral Committee:

Professor Alexander Tsodikov, Chair
Associate Professor Kelley Kidwell
Associate Professor Rafael Meza
Professor Susan Murray

Wenjia Wang
icywang@umich.edu
ORCID iD: 0000-0001-7228-6675

ACKNOWLEDGEMENTS

I would like to express my greatest appreciation to my advisor Professor Alexander Tsodikov for his invaluable guidance, endless support and patience throughout my PhD study. He taught me how to think and be a good statistician. Though he has the most solid theoretical background, every time I had a technical question, he always explained and helped me understand it in a very intuitive way. He is extremely patient with me in guidance of research and is always fully supportive. During this PhD journey, there were hundreds of times I felt frustrated and shamed of myself for little progress. He never blamed me for it, but instead sat down with me to investigate the problem and always had a solution to help me get out of difficulties. I could never have completed this dissertation without his instruction and encouragement.

I am also very grateful to Professor Kelley Kidwell, Susan Murray and Rafeal Meza, for kindly agreeing to be my committee members and contributing insightful suggestions and constructive inputs toward the dissertation. As my former GSRA advisor, Professor Susan Murray also provided support and guidance in my early PhD study. Working with her and collaborators in the pulmonary division in the University of Michigan Hospital makes me equipped with the scientific thinking from the perspective of clinical and biological consideration.

Many thanks to all the faculty and staff in the Biostatistics Department for the guidance and assistance, and to all my friends for our wonderful moments together.

Last but not least, special thanks to my parents, Xingping Wang and Xiaoling Yu, for their unlimited love, belief and support along these years.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (Stanford et al., 1999) was conducted in the 1990s to evaluate the screening effect of prostate-specific antigen (PSA) tests on prostate cancer. While PSA tests led to increased incidence of cancer diagnosis, their effect on prostate cancer mortality is disputed. Treatment for prostate cancer is confounded by indication provided by the cancer diagnosis that can occur earlier with PSA screening.

The null hypothesis of no benefit of screening implies unchanged prostate cancer mortality despite a profound effect of screening on the timing of treatment occurring at diagnosis. The difficulty in designing models to test the null hypothesis lies in the general positive dependence induced by diagnosis always preceding cancer death. Traditional models incorporating this dependence usually let an effect on the intermediate event (diagnosis) propagate through the dependence structure to the terminal event (death), thus excluding the null hypothesis.

This work focuses on developing semiparametric joint models incorporating the null hypothesis. Chapter II proposes a parametric conditional model for cancer incidence given death. We use nonparametric maximum likelihood to develop estimators and establish inference procedures and large sample properties. We conduct simulation studies to illustrate the finite-sample properties of the method; its use in practice is demonstrated with the analysis of the PLCO prostate cancer data, combined with the uncontaminated controls simulated using SEER data before PSA.

In Chapter III, we extend the proposed parametric conditional model of Chapter II to a semiparametric model based on a multinomial logit model for the incident

event. To overcome the computational difficulty from the complexity of a multinomial likelihood function, an artificial latent variable is introduced to transform the multinomial likelihood to Poisson-type, and an EM algorithm, treating the artificial variable as missing data, is derived to obtain the nonparametric maximum likelihood estimators. To illustrate our method, we study its performance in simulations and apply it to the prostate cancer data.

In Chapter IV, we explore a mechanistic joint model to study the problem and develop a test for the causal effect of screening. Applying the concept of modulation that describes how the occurrence of a latent event affects the risk of a future event in a time-dependent fashion, the proposed model belongs to a class of stochastic process frailty models. The profile likelihood-based method is considered for inference. We study efficiency gains of joint modeling recognizing a common cancer progression process driving incidence and mortality. The method is illustrated with simulations and analysis of the prostate cancer data.

CHAPTER I

Introduction

Prostate cancer is one of the most common types of cancer in American men, second to skin cancer (American Cancer Society, 2020). About 1 man in 8 will be diagnosed with prostate cancer during his lifetime. It is also the second leading cause of cancer death for American men, behind only lung cancer. Though prostate cancer can be a serious disease, most men diagnosed with prostate cancer do not die from it. In fact, about 1 man in 41 will die of prostate cancer. The Surveillance, Epidemiology, and End Results (SEER) registry tracks 5-year relative survival rates for prostate cancer in the United States, based on the disease stage at diagnosis (localized, regional and distant). For those diagnosed with prostate cancer between 2010 and 2016, 5-year relative survival rate is almost 100% for patients diagnosed at the localized/regional stage, while it is only 30% for those diagnosed at the distant stage (SEER, 2020). Thus, to prolong cancer survival, early diagnosis is important.

Screening in the population can bring life-saving benefit via early diagnosis and early treatment, yet it may also cause overdiagnosis and overtreatment (Andriole et al., 2009). It is thus crucial to assess the balance between benefits and harms associated with screening. Specifically, for prostate cancer, the screening tool is prostate-specific-antigen (PSA) testing, which entered clinical practice in 1988. Numerous observational studies have reported conflicting findings regarding the benefits

of screening (Lin et al., 2008). A nation-wide randomized, controlled trial of prostate-cancer screening was conducted in the U.S. - the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. For the prostate component of the PLCO trial, from 1993 through 2001, 76,693 men at 10 U.S. study centers were randomly assigned to receive either annual screening or usual care as the control (PLCO Trial, 2010). Men in the screening group were offered annual PSA testing for 6 years. Once diagnosed, the distributions of treatment given were similar in the two groups within each tumor stage. The subjects were followed up for a maximum of 13 years. Researchers found that while PSA tests led to increased incidence of cancer diagnosis, their effect on prostate cancer mortality was in doubt (Andriole et al., 2012). In 2012, the U.S. Preventive Services Task Force (USPSTF) determined that there was "very low probability of preventing a death from prostate cancer in the long term" and recommended against routine use of the test (Force et al., 2008).

It is the purpose of this dissertation to assess the screening effect on cancer mortality. Generally, screening tests affect both cancer diagnosis and death through the following mechanism:

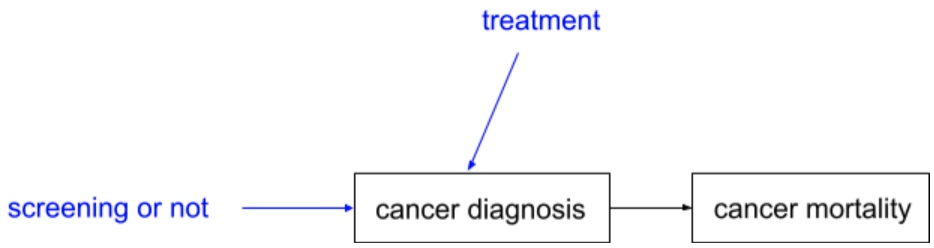

Figure 1.1: Diagram illustrating the disease progression process with screening tests.

1. Cancer diagnosis may occur earlier at potentially earlier stages with screening tests.

2. Assume once diagnosed, specific appropriate treatment will be given, and treatment is confounded by indication provided by the cancer diagnosis.

3. Cancer severity at diagnosis and the following treatment received will affect cancer mortality.

Mathematically speaking, we will test the null hypothesis H_0: NO screening benefit on cancer mortality, despite a profound screening effect on the timing of treatment occurring at diagnosis. Here, intermediate (cancer diagnosis) and terminal (cancer death) events are driven from a common disease progression process, thus holding a strong relationship. A joint modeling approach can be adopted to model both events simultaneously while considering this relationship, which characterizes the disease natural history. Moreover, its predictive capability of prognosis and survival rate provides important insights for optimal treatment decisions. Therefore, we test the screening effect on cancer mortality with joint modeling of cancer diagnosis and death.

The key to establishing a joint model is to characterize the relationship between intermediate and terminal events. Dependent on this relationship, a number of methods have been proposed.

- Semicompeting risk: It denotes the mechanism where the terminal event may censor the intermediate event, but not vice versa. The correlation between events can be captured through copulas (Clayton, 1978; Fine et al., 2001; Chen, 2012) and illness-death models (Andersen et al., 1991, 2012; Xu et al., 2010; Kalbfleisch and Prentice, 2011). The reason why these methods do not work in our scenario is that the intermediate and terminal events are unordered. For example, the terminal event is death due to metastases; while the intermediate event is defined as local recurrence, which does not necessarily precede the terminal.

- Recurrent events: It denotes the mechanism where intermediate always precedes terminal event, which satisfies the time ordering of the events. Gap time analysis (Lin et al., 1999; Huang and Liu, 2007; Shu and Schaubel, 2016) is

a typical analysis method. It jointly models time-to-intermediate and the gap time between time-to-intermediate and time-to-terminal. Yet our null hypothesis is to test the screening effect on the marginal terminal event, which is not directly modeled with gap time analysis. Thus, this method is not compatible with our null hypothesis. Alternatively, the order restriction can be imposed by a conditional argument restricting the domain of an unordered joint distribution of time-to-intermediate and time-to-terminal to the upper wedge. However, with this model, due to positive dependence between intermediate and terminal events, we cannot make the covariate affect one event while not affecting the other, which excludes our null hypothesis.

The difficulty in designing joint models to test the null hypothesis lies in the general positive dependence induced by diagnosis always preceding death. Traditional joint models dealing with this dependence usually let an effect on the diagnosis propagate through the dependence structure to the terminal event, thus excluding the null hypothesis. Throughout this dissertation, we focus on developing semiparametric joint models incorporating the null hypothesis.

In Chapter II, the proposed joint model is composed of two parts; one is to characterize the marginal death distribution, and the other is for conditional incidence of diagnosis given death. The marginal death model enables us to test the screening effect in a straightforward way; while for the conditional model, due to the time ordering of the two events, the parametric beta distribution is a natural choice to provide limited support for the incidence event. Nonparametric maximum likelihood estimation based on semiparametric regression analysis is used for statistical inference, and asymptotic properties of the proposed estimators are studied using empirical process and martingale arguments. The methodology is illustrated with simulations and the PLCO prostate cancer data, combined with uncontaminated controls simulated from SEER data before PSA.

In Chapter III, we extend the proposed parametric conditional model of Chapter II to a semiparametric model based on a multinomial logit model for the incidence event. The technical difficulty comes from the complex form of a multinomial likelihood. EM algorithm can be viewed as a way to replace maximization of the original observed likelihood by the corresponding complete-data likelihood, which may offer computational advantage and increased stability. With this idea in mind, an artificial latent variable is introduced to transform the multinomial likelihood to Poisson-type (Baker, 1994; Lang, 1996; Tsodikov and Chefo, 2008), and an EM algorithm, treating the artificial variable as missing data, is derived to simplify the maximization step to obtain the maximum likelihood estimators (MLE). To illustrate our method, we study its performance with Monte Carlo simulations and real data analysis.

In Chapter IV, we consider the situation where both incidence of cancer diagnosis and death are observed, yet the type of incidence (causal to death or not) is missing. Overdiagnosis is caused if non-causal incidence is detected. This mechanism thus incorporates the null hypothesis, where screening does not affect cancer mortality, though it increases the risk of the incidence event. A mechanistic joint model is formulated on the partially observed disease progression process, and a test is developed for the causal effect of screening. We apply the concept of modulation that describes how the occurrence of a latent event affects the risk of a future event from the same cancer progression process in a time-dependent fashion (Cook and Lawless, 2007). Such formulation results in a strong link and information sharing between cancer incidence and death, thus providing efficiency gains. A likelihood-based approach is adopted for statistical inference, and the obtained MLEs are proved to be consistent and asymptotically efficient. The method is illustrated with numerical examples.

Before starting the next chapter, we will now briefly review some concepts of lead time and length bias in the context of screening.

Lead time (Gates, 2001) is the amount of time by which a diagnosis has been

advanced by screening. In the analysis of survival from diagnosis, lead time constitutes an artificial addition to the survival time for screen-detected cases. To avoid lead time bias, the starting point for survival analysis should be taken as the randomization time, instead of the diagnosis time.

Length bias (Zelen and Feinleib, 1969) is a form of selection bias, where the diagnostic screen does not detect individuals at random, but detects those with slow-growing pre-clinical diseases, due to their longer pre-clinical sojourn time. The individuals picked out by screening will have longer survival, regardless of whether there is a gain in survival due to treating the disease earlier. The extreme form of length bias is overdiagnosis, which would never have caused any symptoms or been clinically detected in the absence of screening. To avoid length bias, when evaluating screening effects from randomized trials, disease-specific survival of all the individuals in the screening group and of all the individuals in the control group should be compared.

CHAPTER II

Conditional Modeling of Incidence of Cancer Diagnosis on Terminal Event

2.1 Introduction

Screening in the population can bring life-saving benefits via early diagnosis and the following early treatment, yet it may also cause overdiagnosis and overtreatment (Andriole et al., 2009). It is thus crucial to assess the balance between benefits and potential harms associated with screening. Specifically, for prostate cancer, the screening tool is prostate-specific-antigen (PSA) testing, which has been widely used in the United States since the late 1980s. Large randomized trials have been conducted to study the screening effect, including the PLCO (Prostate, Lung, Colorectal, and Ovarian) and the ERSPC (European Randomized Screening Study of Screening for Prostate Cancer) trials. Researchers found that while PSA tests led to increased incidence of cancer diagnosis, their effect on prostate cancer mortality was in doubt (Andriole et al., 2012). In 2012, the U.S. Preventive Services Task Force (USPSTF) determined that there was "very low probability of preventing a death from prostate cancer in the long term" and recommended against routine use of the tests (Force et al., 2008). In this article, we are interested in evaluating whether screening can bring benefit on cancer mortality, despite a profound effect on the timing of treatment

occurring at diagnosis.

Motivated by the idea that there is a common process driving the intermediate (incidence of cancer diagnosis) and the terminal (cancer death) events, a joint modeling approach is needed to capture the relationship between the two events. Semicompeting risks (Fine et al., 2001) can model events where the terminal event may censor the intermediate one, but not vice versa (e.g, metastasis and death from local disease); yet fail to model sequential events, as in our scenario. Multi-state models (Andersen et al., 1991, 2012; Kalbfleisch and Prentice, 2011) use transition intensities to model the process where subjects move from one state to the next, thus depicting the event history. This class of models can be easily adapted to different data structures. Xu et al. (2010) proposed an illness-death model with a shared frailty to model the semicompeting risks data. The multi-state model proposed by Dejardin et al. (2010) with two events, progression and death, assumed an ordering to the two events. Yet, due to the positive dependence between intermediate and terminal events, we cannot make the covariate affect one event while not affecting the other with this model, which excludes our null hypothesis. Gap time analysis (Lin et al., 1999; Schaubel and Cai, 2004; Huang and Liu, 2007; Shu and Schaubel, 2016) is also a typical method to analyze data with or without ordering. It jointly models the time to the intermediate event and the gap time between time-to-intermediate and time-to-terminal events. Yet our null hypothesis is to assess the covariate effect on the terminal event, which is not directly modeled with gap time analysis.

Kong et al. (2018) conditionally modeled longitudinal data with terminal event to provide an intuitive and meaningful interpretation of the effect of the terminal event on the longitudinal measures in the joint analysis. Motivated by this idea, we propose a semiparametric joint model with two parts: one is to characterize the marginal death distribution, and the other is for conditional incidence of diagnosis given death. The marginal death model enables us to test the screening effect in an

explicit way, and the conditional model can provide limited support for the incidence event, such that incidence always precedes death. Statistical inference is based on nonparametric maximum likelihood estimation.

The rest of the chapter is organized as follows. Section 2.2 describes the framework of our proposed model and derives its essential distributional characteristics. Section 2.3 presents the likelihood in counting process form and the corresponding martingale properties, as well as the estimating procedure for the nonparametric maximum likelihood estimators. The asymptotic properties are outlined in Section 2.4, with proofs given in the Appendix A.5. Section 2.5 conducts simulation studies, evaluating the performance of the proposed estimators with finite samples. Section 2.6 analyzes the prostate cancer data. Finally, we discuss the results in Section 2.7.

2.2 Model and Likelihood

2.2.1 Data Structure and Notation

Consider two sequential events, such as incidence of cancer diagnosis and cancer death, in our model. Let T_I and T_D be the time-to-incidence and time-to-death, respectively; $\mathbf{z}$ be a set of fully observed covariates; and C be the censoring time which is independent of T_I and T_D given $\mathbf{z}$. For the sequential events, the underlying assumption is incidence must precede the terminal event, i.e. $T_D > T_I \ \ w.p.\ 1$.

We observe $(X_1, \Delta_1, X_2, \Delta_2)$, where $X_1 = min(T_I, C)$ is the time to the intermediate event (i.e. incidence or censoring); $X_2 = min(T_D, C)$ is the time to the terminal event (i.e. death or censoring); $\Delta_1 = \mathbb{1}(X_1 = T_I)$ is the indicator of observing incidence; and $\Delta_2 = \mathbb{1}(X_2 = T_D)$ is the indicator of observing death.

2.2.2 Model Specification

We formulate our model in two parts. The first is to characterize marginal distribution of death with a Cox proportional hazards model, and the second is to model the conditional distribution of incidence given death via beta distribution:

$$d\Lambda_{T_D}(t|\mathbf{z}) = \eta(\mathbf{z})dH(t), \tag{2.1}$$

$$f_{T_I|T_D}(t|t_D) = \frac{1}{B\left(a(\mathbf{z}), b(\mathbf{z})\right)} \left(\frac{t}{t_D}\right)^{a(\mathbf{z})-1} \left(1 - \frac{t}{t_D}\right)^{b(\mathbf{z})-1} \frac{1}{t_D}, \quad 0 < t < t_D. \tag{2.2}$$

Here, $B(a(\mathbf{z}), b(\mathbf{z}))$ is a beta function parameterized by $a(\mathbf{z})$ and $b(\mathbf{z})$. Covariates $\mathbf{z}$ enter the model through η, a and b. Specifically, we have $\eta(\mathbf{z}) = e^{\beta_\eta \mathbf{z}}$, $a(\mathbf{z}) = e^{\beta_a \mathbf{z}}$, $b(\mathbf{z}) = e^{\beta_b \mathbf{z}}$, and $\beta = (\beta_\eta, \beta_a, \beta_b)$ is the combined vector of regression coefficients. The cumulative baseline hazard $H(t)$ summarizes the underlying disease progression pattern leading to a terminal event.

Beta distribution is a natural choice to provide limited support for $T_I \in (0, T_D)$, which ensures that incidence always precedes death. The proposed model directly specifies marginal death, which incorporates our null hypothesis and enables us to test the screening effect on marginal death in a straightforward way.

It may be observed that if T_D goes to infinity, the conditional distribution of incidence becomes 0, which restricts the support of T_D to be finite. Suppose t_{lf} is the last observed failure of death. We set a threshold τ, and assume that for individuals with terminal events censored after t_{lf}, they live up to time τ. Note τ is the time specified after t_{lf}.

2.2.3 Likelihood Construction

We combine (2.1) and (2.2) to build the joint likelihood of cancer diagnosis incidence and death. The likelihood of a single subject with observed data $(X_1, \Delta_1, X_2, \Delta_2)$ falls into one of the following scenarios:

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$):

$$\begin{aligned} L_{11} &= f_{T_D}(X_2) f_{T_I|T_D}(X_1|X_2) \\ &= \eta dH(X_2) e^{-\eta H(X_2)} \frac{1}{B(a,b)} \left(\frac{X_1}{X_2}\right)^{a-1} \left(1 - \frac{X_1}{X_2}\right)^{b-1} \frac{1}{X_2}. \end{aligned}$$

Denote $f_{beta}(X_1; X_2, a, b) = \dfrac{1}{B(a,b)} \left(\dfrac{X_1}{X_2}\right)^{a-1} \left(1 - \dfrac{X_1}{X_2}\right)^{b-1} \dfrac{1}{X_2}$, then

$$L_{11} = \eta dH(X_2) e^{-\eta H(X_2)} f_{beta}(X_1; X_2, a, b).$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 0$):

   - If $X_2 > t_{lf}$,

$$\begin{aligned} L_{10} &= f_{T_D}(\tau) f_{T_I|T_D}(X_1|\tau) \\ &= e^{-\eta H(t_{lf})} \frac{1}{B(a,b)} \left(\frac{X_1}{\tau}\right)^{a-1} \left(1 - \frac{X_1}{\tau}\right)^{b-1} \frac{1}{\tau} \\ &= e^{-\eta H(t_{lf})} f_{beta}(X_1; \tau, a, b). \end{aligned}$$

   - If $X_2 \leq t_{lf}$,

$$\begin{aligned} L_{10} &= \int_{X_2}^{t_{lf}} f_{T_D}(t_D) f_{T_I|T_D}(X_1|t_D)\, dt_D + f_{T_D}(\tau) f_{T_I|T_D}(X_1|\tau) \\ &= \int_{X_2}^{t_{lf}} \eta e^{-\eta H(t_D)} f_{beta}(X_1; t_D, a, b) dH(t_D) + e^{-\eta H(t_{lf})} f_{beta}(X_1; \tau, a, b). \end{aligned}$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = \Delta_2 = 0$):

- If $X_2 > t_{lf}$,

$$L_{00} = \int_{X_2}^{\tau} f_{T_D}(\tau) f_{T_I|T_D}(t_I|\tau)\, dt_I$$
$$= e^{-\eta H(t_{lf})} \int_{\frac{X_2}{\tau}}^{1} \frac{1}{B(a,b)} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} d\left(\frac{t_I}{\tau}\right).$$

Denote the regularized incomplete beta function $I_x(a,b) = \int_0^x \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1} dt$, then

$$L_{00} = e^{-\eta H(t_{lf})} \big[1 - I_{\frac{X_2}{\tau}}(a,b)\big].$$

- If $X_2 \leq t_{lf}$,

$$L_{00} = \int_{X_2}^{t_{lf}} \int_{X_2}^{t_D} f_{T_D}(t_D) f_{T_I|T_D}(t_I|t_D)\, dt_I dt_D + \int_{X_2}^{\tau} f_{T_D}(\tau) f_{T_I|T_D}(t_I|\tau)\, dt_I$$
$$= \int_{X_2}^{t_{lf}} \eta e^{-\eta H(t_D)} \big[1 - I_{\frac{X_2}{t_D}}(a,b)\big] dH(t_D) + e^{-\eta H(t_{lf})} \big[1 - I_{\frac{X_2}{\tau}}(a,b)\big].$$

We can alternatively express the joint log-likelihood for a single subject as:

$$\begin{aligned} l &= \Delta_1\Delta_2 \log\big[f_{T_I}(X_1) f_{T_D|T_I}(X_2|X_1)\big] + \Delta_1(1-\Delta_2)\log\big[f_{T_I}(X_1) S_{T_D|T_I}(X_2|T_I = X_1)\big] \\ &+ (1-\Delta_1)\log S_{T_I}(X_1) \\ &= \Delta_1 \log f_{T_I}(X_1) + (1-\Delta_1)\log S_{T_I}(X_1) \\ &+ \Delta_1\big[\Delta_2 \log f_{T_D|T_I}(X_2|X_1) + (1-\Delta_2)\log S_{T_D|T_I}(X_2|T_I = X_1)\big]. \end{aligned} \tag{2.3}$$

The joint log-likelihood can be partitioned into two parts, such that the contribution from incidence is separated from terminal event. If we can denote l_1 and l_2 as the quantities in each line of equation (2.3), it is easy to see that l_1 is based on information

from incidence, while l_2 is based on additional information from the subsequent time segment between incidence and death.

2.2.4 Prediction of Death Given Incidence

The model also allows us to make predictions of the distribution of the time to the terminal event, given observed incidence information. This is of particular interest to clinical practice, as it allows us to predict survival for a subject who has/ has not been diagnosed after some specified time t^*. Derived in Appendix A.1 is the predicted conditional survival functions for death given patient's diagnosis information. Specifically, we have the survival functions

$$
\begin{aligned}
&Pr(T_D > t | T_I = t^*) \\
&= \begin{cases} \dfrac{\int_t^{t_{lf}} \eta e^{-\eta H(t_D)} f_{beta}(t^*; t_D, a, b) dH(t_D) + e^{-\eta H(t_{lf})} f_{beta}(t^*; \tau, a, b)}{\int_{t^*}^{t_{lf}} \eta e^{-\eta H(t_D)} f_{beta}(t^*; t_D, a, b) dH(t_D) + e^{-\eta H(t_{lf})} f_{beta}(t^*; \tau, a, b)}, & t \geq t^* \\ 1, & t < t^* \end{cases}
\end{aligned} \tag{2.4}
$$

for a subject who has been diagnosed at t^*, and

$$
\begin{aligned}
&Pr(T_D > t | T_I > t^*) \\
&= \begin{cases} \dfrac{\int_t^{t_{lf}} \eta e^{-\eta H(t_D)} \left[1 - I_{\frac{t^*}{t_D}}(a, b)\right] dH(t_D) + e^{-\eta H(t_{lf})} \left[1 - I_{\frac{t^*}{\tau}}(a, b)\right]}{\int_{t^*}^{t_{lf}} \eta e^{-\eta H(t_D)} \left[1 - I_{\frac{t^*}{t_D}}(a, b)\right] dH(t_D) + e^{-\eta H(t_{lf})} \left[1 - I_{\frac{t^*}{\tau}}(a, b)\right]}, & t \geq t^* \\ 1, & t < t^* \end{cases}
\end{aligned} \tag{2.5}
$$

for a subject who has not been diagnosed until t^*.

2.3 Estimation

The proposed model is semiparametric, consisting of a parametric component β for covariate effects and nonparametric components $H(\cdot)$ for baseline hazard of death.

$H(\cdot)$ is a non-decreasing step function with jumps $\{dH\}$ only at the observed death times. Let us denote the full parameter set $\Omega = (\beta, \{dH\})$. We use the EM algorithm (Tsodikov (2003), Rice and Tsodikov (2017)) and profile likelihood (Murphy and Van der Vaart (2000)) approach to obtain the nonparametric maximum likelihood estimator (NPMLE) for Ω.

2.3.1 Martingale Theory

In counting process notation, for subject i, let $N_{1i}(t) = \mathbb{1}(X_{1i} \leq t, \Delta_{1i} = 1)$ and $Y_{1i}(t) = \mathbb{1}(X_{1i} \geq t)$ be the observed counting process and at risk process for incidence, respectively; let $N_{2i}(t) = \mathbb{1}(X_{2i} \leq t, \Delta_{2i} = 1)$ and $Y_{2i}(t) = \mathbb{1}(X_{2i} \geq t)$ denote the observed counting process and at risk process for death. Log-likelihood (2.3) can be rewritten in counting process form as

$$l = \sum_i l_i = \sum_i (l_{1i} + l_{2i}), \tag{2.6}$$

where

$$\begin{aligned} l_{1i} &= \int_0^\tau \log d\Lambda_{1i}(t) dN_{1i}(t) - Y_{1i}(t) d\Lambda_{1i}(t), \\ l_{2i} &= \int_0^\tau \Big[\int_{t_1}^\tau \log d\Lambda_{2i}(t|t_1) dN_{2i}(t) - Y_{2i}(t) d\Lambda_{2i}(t|t_1) \Big] dN_{1i}(t_1). \end{aligned}$$

The martingales $dM_{1i}(t)$ and $dM_{2i}(t|t_1)$ can be constructed based on observed counting processes with respect to filtration $\mathcal{F}_i(t-) = \sigma\big\{N_{1i}(s), N_{2i}(s), Y_{1i}(s), Y_{2i}(s), \boldsymbol{z}_i : s \in [0,t)\big\}$ as

$$\begin{aligned} dM_{1i}(t) &= dN_{1i}(t) - Y_{1i}(t) d\Lambda_{1i}(t), \\ dM_{2i}(t|t_1) &= dN_{2i}(t) dN_{1i}(t_1) - Y_{2i}(t) dN_{1i}(t_1) d\Lambda_{2i}(t|t_1). \end{aligned}$$

Here,

$$d\Lambda_{1i}(t) = \frac{\int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} f_{beta}(t; t_D, a_i, b_i) dH(t_D) + e^{-\eta_i H(t_{lf})} f_{beta}(t; \tau, a_i, b_i)}{\int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} \big[1 - I_{\frac{t}{t_D}}(a_i, b_i)\big] dH(t_D) + e^{-\eta_i H(t_{lf})} \big[1 - I_{\frac{t}{\tau}}(a_i, b_i)\big]}$$

$$\triangleq \Theta_{1i}(t),$$

$$d\Lambda_{2i}(t|t_1) = \frac{\eta_i e^{-\eta_i H(t)} f_{beta}(t_1; t, a_i, b_i)}{\int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} f_{beta}(t_1; t_D, a_i, b_i) dH(t_D) + e^{-\eta_i H(t_{lf})} f_{beta}(t_1; \tau, a_i, b_i)} dH(t)$$

$$\triangleq \Theta_{2i}(t; t_1) dH(t).$$

$d\Lambda_{1i}(t)$ is the hazard of subject i having incidence at time t, and $d\Lambda_{2i}(t|t_1)$ is the hazard of subject i dying of cancer at time t, given incidence at time t_1. They can be derived through the following probabilistic argument:

$$\mathbb{E}\{dN_{1i}(t)|\mathcal{F}_i(t-)\} = Y_{1i}(t) Pr\{dN_{1i}(t) = 1 | Y_{1i}(t) = 1\} = Y_{Ii}(t) d\Lambda_{1i}(t),$$

and $\mathbb{E}\{dN_{2i}(t|t_1)|\mathcal{F}_i(t-)\} = Y_{2i}(t) Pr\{dN_{2i}(t|t_1) = 1 | Y_{2i}(t|t_1) = 1\} = Y_{2i}(t) d\Lambda_{2i}(t|t_1)$ (see Appendix A.2 for more details).

2.3.2 Score Functions and NPMLE

Define partial derivatives of $\Theta_{1i}(t)$ and $\Theta_{2i}(t;t_1)$, with respect to $\{dH(s)\}$ and β, respectively, as

$$\begin{aligned}
\dot{\Theta}_{1i,dH(s)}(t) &= \frac{\partial \Theta_{1i}(t)}{\partial dH(s)}, \\
\dot{\Theta}_{1i,\beta}(t) &= \frac{\partial \Theta_{1i}(t)}{\partial \beta}, \\
\dot{\Theta}_{2i,dH(s)}(t;t_1) &= \frac{\partial \Theta_{2i}(t;t_1)}{\partial dH(s)}, \\
\dot{\Theta}_{2i,\beta}(t;t_1) &= \frac{\partial \Theta_{2i}(t;t_1)}{\partial \beta}.
\end{aligned}$$

As in Hu and Tsodikov (2013), for a functional $J(f)$, $f = f(x)$, the functional derivative in the above equations is defined as

$$\frac{\partial J(f)}{\partial df(s)} = \left.\frac{\partial J(f+\epsilon g)}{\partial \epsilon}\right|_{\epsilon=0, g=\mathbb{1}(x>s)}.$$

Applying the functional derivative to the full log-likelihood (2.6), with respect to the infinite-dimensional parameters $\{dH(s)\}$, we can obtain the score function for $\{dH(s)\}$ as (see Appendix A.3 for details)

$$\begin{aligned}
U_{dH(s)} = \sum_i \Bigg\{ & \int_s^\tau \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t) \\
& + \int_s^\tau \int_0^t \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) + \int_0^s \frac{dM_{2i}(s;t_1)}{dH(s)} \Bigg\},
\end{aligned} \tag{2.7}$$

which is a martingale under the true model.

Taking derivative of the log-likelihood, with respect to the regression parameter

β, we can have the score function for β as

$$U_\beta = \sum_i \left\{ \int_0^\tau \frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_0^\tau \int_0^t \frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \right\}, \tag{2.8}$$

which is also a martingale under the true model.

Set score functions (2.7) and (2.8) to be zero, and solve them, can give the NPMLE $\hat{\Omega} = (\hat{\beta}, \{d\hat{H}\})$.

2.3.3 Estimation Procedure - EM Algorithm

The NPMLE can be obtained using $(\hat{\beta}, \{d\hat{H}\}) = \text{argmax}_{\beta,\{dH\}}\, l(\beta, \{dH\})$. However, it is unpleasant to directly maximize, since the nonparametric parts $\{dH\}$ are of infinite-dimension. Instead, we apply the profile likelihood to estimate β and $\{dH\}$ jointly. We first obtain the estimators of $\{dH\}$ with fixed β. Then replacing $\{dH\}$ in the observed log-likelihood $l(\beta, \{dH\})$ with $\{d\hat{H}(\beta)\}$, we have the profile log-likelihood $l_{pr} = l\big(\beta, \{d\hat{H}(\beta)\}\big)$. Finally, the finite-dimensional parameter β is estimated by maximizing the resulting profile likelihood over β.

The key step is to obtain the estimator of $\{dH\}$, given β. We use EM algorithm to estimate the baseline hazards. Derivation of EM algorithm for our model is shown in Appendix A.4. It gives us the score functions for $\{dH\}$ in $(k+1)th$ iteration as

$$U_{dH^{(j+1)}}(s) = \sum_{i=1}^n \left\{ \frac{dN_{2,i}(s)}{dH^{(j+1)}(s)} - \Psi_i^{(j)}(s) + \Big[\frac{dH^{(j)}(s)}{dH^{(j+1)}(s)} - 1\Big]\theta_i^{(j)}(s) \right\} = 0, \tag{2.9}$$

where

$$\Psi(s) = \mathbf{1}(X_2 \le t_{lf}) \Big\{ Y_2(s)\eta + [1 - Y_2(s)](1-\Delta_2)\Big[\Delta_1 \frac{\eta U(s^-, t_{lf}) + \eta V - U.s}{U(X_2, t_{lf}) + V} + (1-\Delta_1)\frac{\eta W(s^-, t_{lf}) + \eta Z - W.s}{W(X_2, t_{lf}) + Z}\Big] \Big\} + \mathbf{1}(X_2 > t_{lf})\eta,$$

$$\theta(s) = \mathbf{1}(X_2 \leq t_{lf})[1 - Y_2(s)](1 - \Delta_2)\left\{\Delta_1 \frac{U.s}{U(X_2, t_{lf}) + V} + (1 - \Delta_1)\frac{W.s}{W(X_2, t_{lf}) + Z}\right\}$$

Here,

$$U.s = \frac{1}{B(a,b)}\eta e^{-\eta H(s)}\left(\frac{X_1}{s}\right)^{a-1}\left(1 - \frac{X_1}{s}\right)^{b-1}\frac{1}{s},$$

$$U(u,v) = \int_u^v U.s\, dH(s),$$

$$W.s = \eta e^{-\eta H(s)}\int_{X_2}^{s}\frac{1}{B(a,b)}\left(\frac{t_I}{s}\right)^{a-1}\left(1 - \frac{t_I}{s}\right)^{b-1}\frac{1}{s}\, dt_I,$$

$$W(u,v) = \int_u^v W.s\, dH(s),$$

$$V = e^{-\eta H(t_{lf})}\frac{1}{B(a,b)}\left(\frac{X_1}{\tau}\right)^{a-1}\left(1 - \frac{X_1}{\tau}\right)^{b-1}\frac{1}{\tau},$$

$$Z = e^{-\eta H(t_{lf})}\int_{X_2}^{\tau}\frac{1}{B(a,b)}\left(\frac{t_I}{\tau}\right)^{a-1}\left(1 - \frac{t_I}{\tau}\right)^{b-1}\frac{1}{\tau}\, dt_I.$$

Solving the equation, we have Breslow-type estimators

$$dH^{(j+1)}(s) = \frac{\sum_i dN_{2,i}(s) + \big[\sum_i \theta_i^{(j)}(s)\big] dH^{(j)}(s)}{\sum_i \big[\Psi_i^{(j)}(s) + \theta_i^{(j)}(s)\big]}.$$

$dH(s)$ are updated iteratively, $k = 0, 1, 2, ...$, until convergence, i.e., $dH^{(k+1)} \to dH^{(k)}$. Note, at convergence, the second term of (2.9) disappears, leaving the estimating equation the same as that obtained from observed data; also estimators are consistent (Tsodikov (2003)).

The estimation procedure is described as follows:

Start with $\beta^{(0)} = 0$, $j = 0$.

1. Maximize the likelihood over $H(\beta)$, given $\beta = \beta^{(j)}$:

   (a) Set $k = 0$. Initialize $d\hat{H}^{(0)}(s)$ such that all jumps in the baseline hazards

have equal size, respectively.

(b) With β fixed, calculate $d\hat{H}^{(k+1)}(s)$ using equations above.

(c) Repeat step (b) to update $d\hat{H}^{(k+1)}(s)$, until convergence, i.e.

$$\left\| d\hat{H}^{(k+1)}(s) - d\hat{H}^{(k)}(s) \right\|_2 < \epsilon$$

2. Maximize the profile log-likelihood $l_{pr}(\beta) = l(\beta, \{d\hat{H}(\beta)\})$ over β:

(a) Calculate profile log-likelihood $l(\beta, \{d\hat{H}(\beta)\})$.

(b) Find $\beta^{(j+1)}$ by maximizing $l_{pr}(\beta)$ over β, i.e. $\beta^{(j+1)} = \text{argmax}_\beta \, l_{pr}(\beta)$, using numerical optimization method, e.g. Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS).

Iteratively apply steps 1-2 to estimate β, until convergence of $l_{pr}(\beta)$

$$l_{pr}(\beta^{(j+1)}) - l_{pr}(\beta^{(j)}) < \xi.$$

Note the convergence tolerance for the inner loop (EM algorithm used to estimate baseline hazards, given β) should be stricter than that for the outer loop, e.g. $\epsilon = 10^{-6}$, $\xi = 10^{-5}$.

2.4 Asymptotic Properties

We apply the empirical process (Kosorok, 2008; Van Der Vaart and Wellner, 2000) and the theory of martingale structure in counting process to build the asymptotic properties, adapted from previous work (Zeng and Lin, 2007, 2010; Chen, 2009, 2010; Hu and Tsodikov, 2014; Rice and Tsodikov, 2017).

Assuming regularity conditions hold, in the following, Theorems II.1 and II.2 state the consistency and weak convergence results of the NPMLE $\hat{\Omega}$, while Theorem II.3

justifies the use of negative Hessian matrix from profile log-likelihood in variance estimation. Regularity conditions and proofs are provided in Appendix A.5.

Under regularity conditions,

Theorem II.1. *With probability 1: $\hat{\beta}$ converges to β^0; $\hat{H}(t)$ converges to $H^0(t)$ uniformly over the interval $[0,\tau]$, respectively. Here, β^0 and $H^0(t)$ are the true values of $\hat{\beta}$, and $\hat{H}(t)$.*

Theorem II.2. *$n^{1/2}\{\hat{\beta}-\beta^0, \hat{H}(t)-H^0(t)\}$ converges weakly to a zero-mean Gaussian process. In addition, consider a linear functional of $\hat{\Omega}$,*

$$n^{1/2}\Big\{a^T(\hat{\beta}-\beta^0)+\int_0^\tau b(t)d\big(\hat{H}(t)-H^0(t)\big)\Big\},$$

where a is a real vector, $b(t)$ is a function with bounded total variation in $[0,\tau]$, evaluated at the observed death times. Let $\Phi^T=(a^T,\{b(\cdot)\}^T)$. The asymptotic variance-covariance function of the linear functional above can be consistently estimated by $\Phi^T(\mathcal{I}_n)^{-1}\Phi$, where $\mathcal{I}_n$ is the observed information matrix for Ω, that is, $\mathcal{I}_n=-\dfrac{\partial^2 l_n}{\partial\Omega\partial\Omega^T}\Big|_{\Omega=\hat{\Omega}}$, where $l_n=n^{-1}\sum_{i=1}^n l_i$, l_i is subject i's observed log-likelihood defined.

Theorem II.3. *The inverse of the negative Hessian matrix of the profile log-likelihood with respect to β is a consistent estimator of the limiting variance-covariance matrix of $\hat{\beta}$. That is,*

$$\Big(-\frac{\partial^2 l_{pr,n}}{\partial\beta\partial\beta^T}\Big|_{\beta=\hat{\beta}}\Big)^{-1}\xrightarrow{p} Var\big[\sqrt{n}(\hat{\beta}-\beta^0)\big],$$

where $l_{pr,n}=n^{-1}\sum_{i=1}^n l_{pr,i}(\beta)=n^{-1}\sum_{i=1}^n l_i\big(\beta,\{d\hat{H}(\beta)\}\big)$.

2.5 Simulation Studies

This section presents Monte Carlo simulations to illustrate our proposed methodology. The simulation settings were as follows. The true baseline hazard for terminal event was $H(t) = 0.1t^2$. We considered two covariates Z_1 and Z_2, where $Z_1 \sim Bernoulli(0.5)$, and $Z_2 \sim Normal(0, 1)$. Covariates entered the model via $\eta(\mathbf{z}) = e^{\beta_1 z_2}$, $a(\mathbf{z}) = e^{\beta_2 z_1 + \beta_3 z_2}$, and $b(\mathbf{z}) = e^{\beta_4 z_1 + \beta_5 z_2}$, with true parameters $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (1, 0.5, 0.3, 0.2, -0.6)$. Censoring was simulated from the exponential distribution $Exp(0.2)$, yielding 25% intermediate and 45% terminal events censored.

2.5.1 Finite-sample Properties of Parameter Estimates

We conducted simulations to study the finite-sample properties of the parameter estimates obtained. Samples of size 200 and 500 were examined, each took two different values of τ, 1 and 1000 units after the last observed failure time t_{lf}. For each simulation scenario, 1000 data sets were generated. Standard errors were obtained from the numerically evaluated Hessian matrix at the solution.

The simulation results are summarized in Table 2.1. Comparing scenarios with sample size 200 and 500, both with $\tau = t_{lf} + 1$, the proposed estimators appear to have satisfactory performance for the sample sizes considered, with diminishing bias for larger sample size and good coverage probability at the 95% nominal level. With the larger sample size, we see better agreement between asymptotic standard errors and empirical standard deviations, which validates the performance of the variance estimators. Comparing scenarios with same sample size 500, yet different choices of τ ($t_{lf} + 1$ and $t_{lf} + 1000$), the estimators both perform well, in terms of small bias, consistent standard error estimation, and well-maintained 95% coverage probability. Thus, the proposed estimators are robust to different specifications of τ.

Table 2.1: Simulation results using the proposed beta model.

N	τ	β	Truth	Bias	ASE	ESD	95% CP
200	+1	β_1	1	-0.008	0.119	0.117	0.956
		β_2	0.5	0.031	0.149	0.153	0.949
		β_3	0.3	0.003	0.109	0.105	0.936
		β_4	0.2	0.030	0.164	0.159	0.949
		β_5	-0.6	-0.004	0.113	0.111	0.938
500	+1	β_1	1	-0.005	0.073	0.073	0.962
		β_2	0.5	0.012	0.098	0.097	0.956
		β_3	0.3	0.007	0.066	0.065	0.949
		β_4	0.2	0.007	0.099	0.100	0.938
		β_5	-0.6	0.007	0.066	0.069	0.967
500	+1000	β_1	1	-0.007	0.072	0.072	0.942
		β_2	0.5	0.009	0.099	0.097	0.951
		β_3	0.3	0.014	0.069	0.066	0.931
		β_4	0.2	0.001	0.100	0.100	0.942
		β_5	-0.6	0.013	0.067	0.070	0.951

ASE: average of estimated standard errors
ESD: empirical standard deviation based on Monte Carlo estimates
95% CP: 95% coverage probability

2.5.2 Hypothesis Testing

Given the proposed estimators, we constructed the likelihood ratio test (LRT) to test the covariate effect on the terminal event.

H_0: z_1 does not have an effect on the marginal terminal event.

The simulation set-up was the same as before. Here, sample size was chosen to be 500 and $\tau = t_{lf} + 1$. The simulation procedure was as follows:

1. Generate a data set under H_0.
2. Fit the models with and without covariate z_1 entering η, respectively, and obtain the two log-likelihoods.
3. Calculate the LRT statistic, and decide if H_0 is rejected.
4. Repeat steps 1-3 1000 times, and calculate the empirical p-value.

Based on the simulations, the 0.05 significance level is well maintained (0.053).

2.6 Application to Prostate Cancer Screening Trial Data

2.6.1 Data

The proposed method was applied to the motivating setting of the prostate cancer screening trial, testing the screening effect on cancer mortality. The data come from the PLCO trial, with patients entering the trial aged 55-74 years old. The control arm in the PLCO trial was contaminated (Vickers, 2017), as about 50% of PLCO control patients had PSA testing before enrollment, and of the remainder, close to 90% had PSA measured during the trial (Shoag et al., 2016). Researchers described the trial as comparing "opportunistic versus systematic screening" rather than screening versus no screening. Thus, we needed to introduce a set of uncontaminated control data from external data to assess the screening effect. Comparing SEER with PLCO data (Pinsky et al., 2012), there was no unambiguous evidence showing a healthy volunteer effect. Therefore, a simulated subset of SEER data, with diagnosis between 1980 and 1987 before the use of PSA as a screening tool, was created to act as uncontaminated "perfect" controls.

In the combined data set, 76,674 subjects are from PLCO trial (38,335 subjects in the screening arm, and 38,339 subjects in the control arm), and 38,335 subjects are from the SEER control arm. For patients from the PLCO trial, 4418 (11.52%) in the screening arm were diagnosed with prostate cancer, and 145 (0.38%) died of it; while 4036 (10.53%) in the control arm were diagnosed and 142 (0.37%) died of it. For patients in the simulated SEER control arm, 2726 (7.11%) were diagnosed with prostate cancer, and 606 (1.58%) died of it. Maximum follow-up time was 13 years. From the data, we have information about time-to-diagnosis and time-to-death. In addition to arm (screening/control) and trial (PLCO/SEER), we are also interested in studying the age effect, since age is an important risk factor for prostate cancer. To make patients from different trials comparable, the simulated SEER data have the

same age distribution as the PLCO data.

2.6.2 Results

In our proposed method, covariates entered the model via

$$\eta(\mathbf{z}) = e^{\beta_{\eta 1} Age + \beta_{\eta 2} Arm + \beta_{\eta 3} Trial}$$

$$a(\mathbf{z}) = e^{\beta_{a0} + \beta_{a1} Age + \beta_{a2} Arm + \beta_{a3} Trial}$$

$$b(\mathbf{z}) = e^{\beta_{b0} + \beta_{b1} Age + \beta_{b2} Arm + \beta_{b3} Trial}$$

The goodness of fit for the Cox regression model for marginal death is checked in Figure 2.1. From the graphical inspection, there is no pattern for Schoenfeld residuals over time, supporting the proportional hazards assumptions for the covariates age, arm and trial, with respective goodness of fit p-values of 0.72, 0.92 and 0.58, respectively.

Table 2.2 shows the regression coefficient estimates for both the proposed joint model and the Cox model for marginal death. Based on the estimates of our proposed joint model, older patients have greater risk of death (β=0.53, HR=1.69, p-value<0.0001), and screening helps reduce cancer mortality (β=-1.16, HR=0.31, p-value<0.0001). Due to the relatively small number of deaths, which may lead to lack of power to detect any difference between arms, and the contamination of PLCO control arm, there is no significant difference between PLCO screening and PLCO control arms.

Regarding the analysis of time-to-death, theoretically, point estimates of β_η's should be similar for the two models; as the table shows, the model results are as expected. The standard errors for β_η's in the proposed method are slightly smaller than that from the marginal model, since we utilize information from incidence of

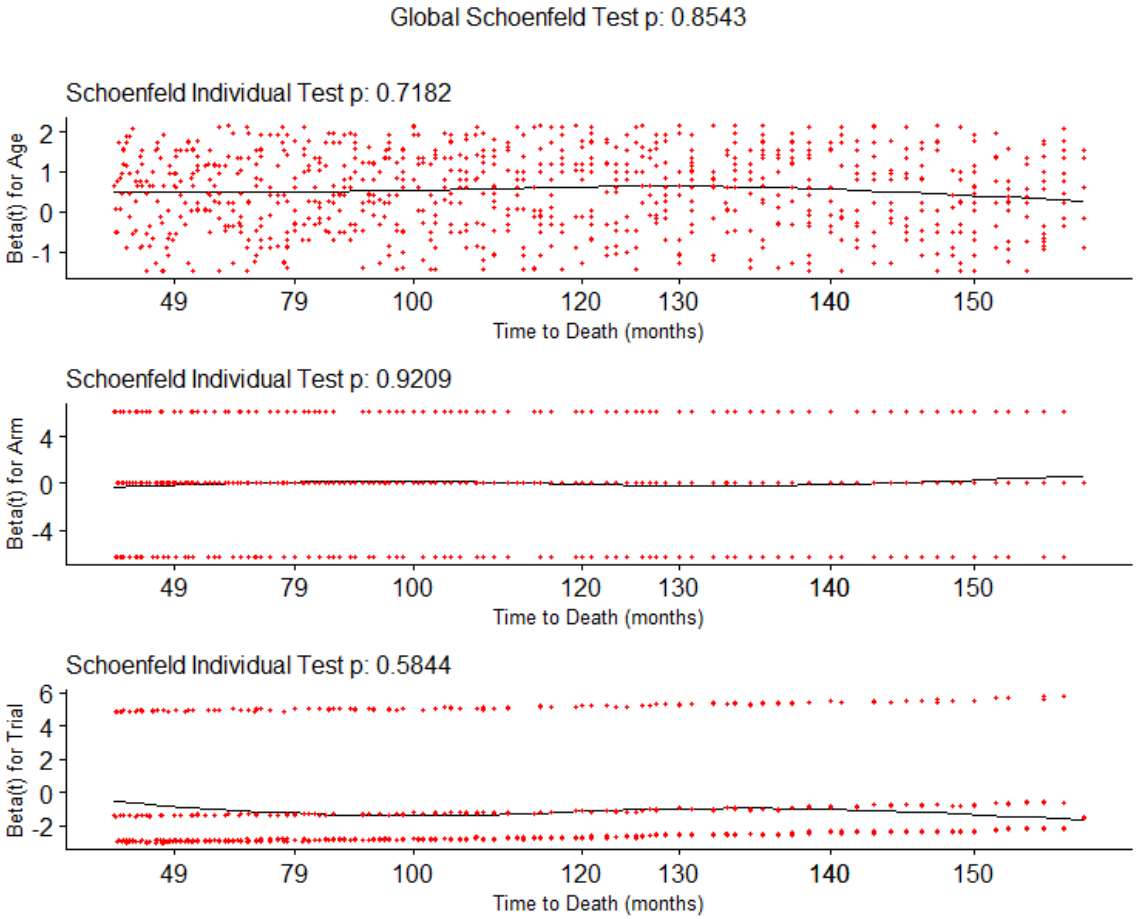

Figure 2.1: Goodness of fit of the Cox regression model for marginal death. The solid line plots the log hazard ratio of the corresponding covariate over time, and the red dots are scaled Schoenfeld residuals.

diagnosis to gain efficiency. Our proposed model allows a bigger picture of disease progression process, as it incorporates incidence data, while the simple Cox model for marginal death does not.

Table 2.2: Prostate cancer screening trial analysis.

(a) Log-hazard ratio for time-to-death (β_η)

	Proposed Joint Model			Cox for Marginal Death		
Parameter	Est.	SE	p-value	Est.	SE	p-value
Age	0.528	0.033	< 0.0001	0.529	0.034	< 0.0001
PLCO scr vs. PLCO control	0.031	0.117	0.791	0.021	0.118	0.861
PLCO control vs. SEER control	-1.155	0.094	< 0.0001	-1.148	0.094	< 0.0001

(b) Parameter estimate for time-to-incidence given death (β_a, β_b)

	Proposed Joint Model		
Parameter	Est.	SE	p-value
β_{a0}	0.607	0.023	< 0.0001
β_{a1}	-0.152	0.011	< 0.0001
β_{a2}	-0.269	0.024	< 0.0001
β_{a3}	-0.629	0.028	< 0.0001
β_{b0}	-1.775	0.038	< 0.0001
β_{b1}	0.017	0.017	0.326
β_{b2}	-0.240	0.037	< 0.0001
β_{b3}	-0.188	0.046	< 0.0001

To assess model fit (Dejardin et al., 2010), Figure 2.2 and 2.3 present the survival estimates for time-to-diagnosis and time-to-death, respectively, for subjects who entered the study at age 67, stratified by arm and trial. Both of the figures have plots of marginal survival functions obtained using the proposed model, which closely match the KM estimates obtained from the observed data, confirming the validity of the model.

2.6.3 Prediction of Death Given Diagnosis

We can also use the proposed joint model to predict cancer mortality given diagnosis information, which is of particular interest to clinical practice. Equations (2.4) and (2.5) allow us to calculate this conditional survival function and produce the

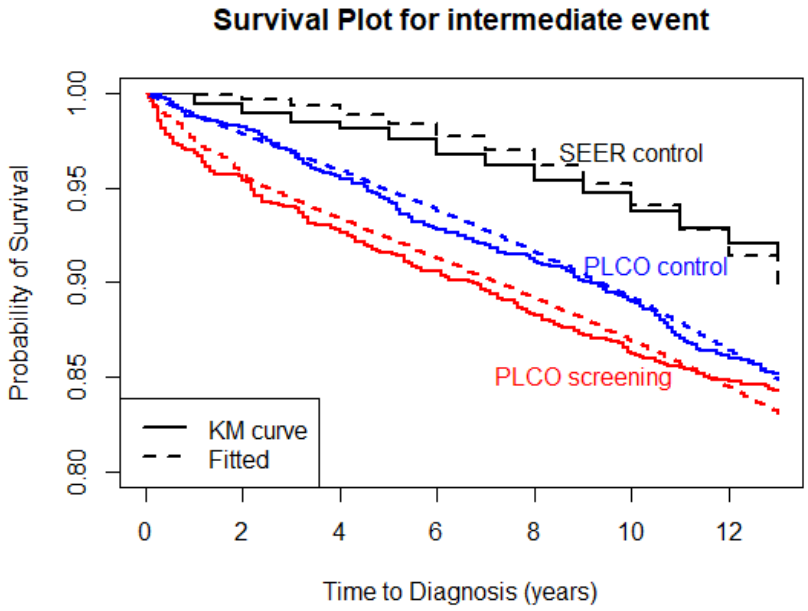

Figure 2.2: Survival estimates for time-to-diagnosis for subjects who entered the study at age 67. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

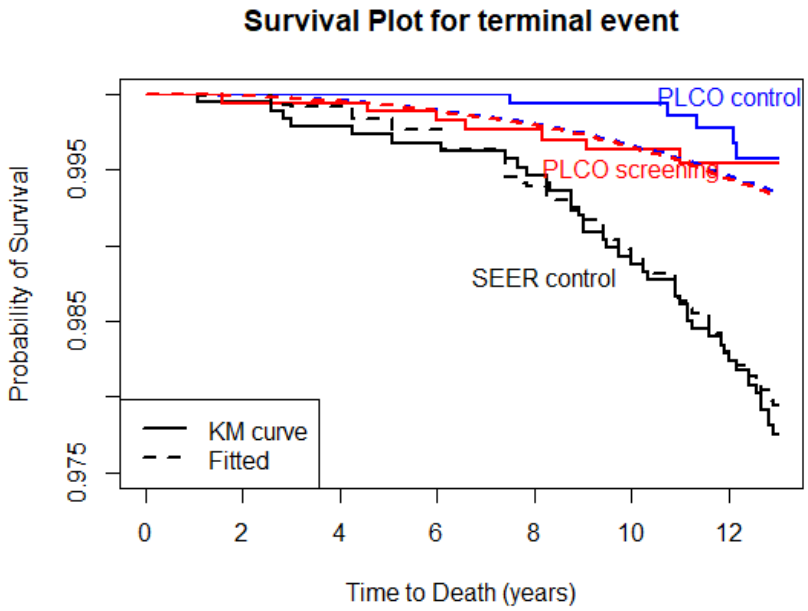

Figure 2.3: Survival estimates for time-to-death for subjects who entered the study at age 67. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

plots shown in Figure 2.4. Figure 2.4 presents the predicted survival rates stratified by arm and trial, given diagnosis status (diagnosed and not diagnosed) after 1 year of randomization, among men who entered the study at age 67. We can see that given the same diagnosis information, patients without screening (SEER control) have much greater risk of death, compared with those with screening (PLCO trial). The PLCO screening arm and PLCO control arm present no surprise and show little difference in predicted survival rates. In addition, note that the survival probabilities for patients with NO cancer diagnosed at 1 year (right plot) are very close to 1, while those with cancer diagnosed (left plot) are not. Thus, in comparison with patients with NO cancer detected, patients with cancer detected are more likely to die from it, as expected.

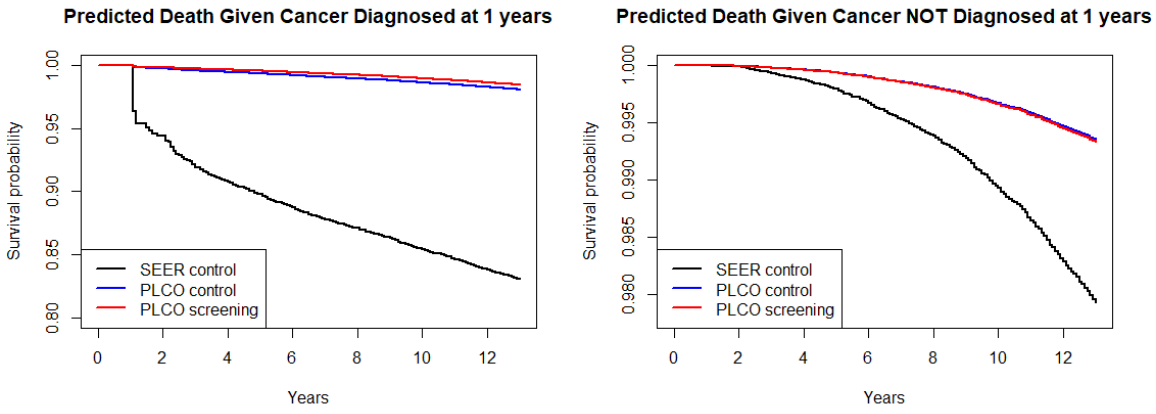

Figure 2.4: Predicted survival estimates for death from prostate cancer, given diagnosis status after 1 year of randomization, among men who entered the study at age 67. Left figure is for a hypothetical subject who has prostate cancer diagnosed at the time indicated, while the right one is for a subject not diagnosed at that time.

2.7 Discussion

We have presented a framework to test the screening effect on cancer mortality with a semiparametric joint model of ordered events. Within the proposed model, the NPMLE can be obtained by EM algorithm and profile likelihood, and its asymptotic

properties established. Simulation studies with different scenarios indicate good finite sample performance. Application of the proposed model to the PLCO prostate cancer data, combined with simulated SEER control data, reveals the benefit of screening on reducing cancer mortality.

The traditional joint models usually let an effect on the diagnosis propagate through the dependence structure to the terminal event, thus excluding the null hypothesis. Our proposed model specifies marginal distribution of death directly, thus allowing testing of the screening effect on marginal death in a straightforward way. Although in the model specification, we set a threshold τ such that no death happened after it, this is a typical assumption when asymptotic properties are proved in survival analysis. Moreover, the proposed estimators are robust with different specifications of τ. In fact, we can assign a very large value to τ, and in a practical sense, we can say there is no upper limit for the terminal event.

The beta conditional incidence model provides limited support for the incidence event, such that the joint pdf is zero when death precedes incidence, which is a true reflection of the sequential mechanism. The conditional model can also be specified with other parametric distributions, such as Weibull, or be extended to a semiparametric model to offer more flexibility.

Compared with the model on marginal death, our joint model provides a predictive utility to help with clinical practice. In addition, it has slight efficiency gain, due to the extra information from observations with incidence observed, yet death censored. In future work, it would be of interest to further improve efficiency, by assuming a stronger relationship between the two events.

CHAPTER III

A Semiparametric Joint Model for Estimating the Screening Effect on Cancer Mortality

3.1 Introduction

Prostate-specific-antigen (PSA) testing, a screening tool for prostate cancer, has been widely used in the United States since the late 1980s. Treatment for prostate cancer can be confounded by indication provided by the cancer diagnosis that can occur earlier with PSA screening. While PSA tests led to increased incidence of cancer diagnosis, their effect on prostate cancer mortality is disputed (Andriole et al., 2012).

In this chapter, we test the screening benefit on cancer mortality regardless of a profound effect of screening on the timing of treatment occurring at cancer diagnosis. We extend the proposed parametric conditional incidence model in Chapter II to a semiparametric model to relax the parametric assumptions, thus providing wider applicability and increased robustness. The sequential mechanism of incidence preceding death can be interpreted mathematically as

$$\int_0^{T_D} f_{T_I|T_D}(t|T_D)dt = 1.$$

In discrete time survival analysis, it can be alternatively written as

$$\sum_{k:0<t_k\leq T_D} Pr(T_I \in [t_k, t_k + dt)|T_D) = 1.$$

Here time before death is split into several non-overlapping time intervals, and $[t_k, t_k + dt)$ is the kth interval. The equation indicates that the summation of probabilities of time-to-incidence falling into any time interval before death is 1, which has a similar form as the constraint of a multinomial logit model. Therefore, a multinomial logit model is considered for the incidence event.

The computation in the multinomial logit model is costly, especially when the outcome has a large number of categories as in our case, since it involves high-dimensional integration and maximization. A number of approaches have been developed to overcome the computational difficulties. Baker (1994) proposed to transform the multinomial likelihood to a Poisson likelihood to reduce the complexity of the likelihood function and simplify maximization at the cost of augmenting the model parameters. This approach, however, inflates the model dimension and increases computational difficulties. Breslow and Clayton (1993) considered penalized quasi-likelihood estimation to avoid the complex form of multinomial likelihood. Rabe-Hesketh et al. (2002) applied Gaussian quadrature to numerically approximate multidimensional integrals in the multinomial problem. Tsodikov and Chefo (2008) introduced an artificial latent variable to transform the multinomial likelihood to a Poisson likelihood by means of expectation, and used the self-consistency framework (Tsodikov, 2003) to apply EM algorithm for estimation. Although this approach introduces an artificial variable, it is averaged over in the E step and does not need to be estimated. In addition, the M step takes factorization of the model dimension and thus provides computational advantages.

We extend Tsodikov and Chefo (2008) to model the incidence event given death,

where outcome is a time-to-event variable and number of categories for the outcome is not pre-specified and varies with each subject. Treating the artificial variable and censored observations as missing data, an EM algorithm is derived for parameter estimation. Section 3.2 specifies the proposed model and derives its essential distributional characteristics. Section 3.3 presents the likelihood in counting process form and the corresponding martingale properties, as well as the estimating procedure for the nonparametric maximum likelihood estimators. The asymptotic properties are outlined in Section 3.4. Section 3.5 demonstrates the performance of the proposed estimators with finite samples in simulation studies. Section 3.6 analyzes the prostate cancer data. Discussion follows in Section 3.7.

3.2 Model and Likelihood

3.2.1 Data Structure and Notation

Let T_I and T_D be the time-to-incidence and time-to-death, respectively; $\mathbf{z}$ be a set of fully observed covariates; and C be the censoring time which is independent of T_I and T_D given $\mathbf{z}$. For the sequential events, the underlying assumption is incidence must precede the terminal event, i.e. $T_D > T_I \ \ w.p.\ 1$.

We observe $(X_1, \Delta_1, X_2, \Delta_2)$, where $X_1 = min(T_I, C)$ is the time to the intermediate event (i.e. incidence or censoring); $X_2 = min(T_D, C)$ is the time to the terminal event (i.e. death or censoring); $\Delta_1 = \mathbb{1}(X_1 = T_I)$ is the indicator of observing incidence; and $\Delta_2 = \mathbb{1}(X_2 = T_D)$ is the indicator of observing death. t_{lf1}, t_{lf2} are the last observed failures of incidence and death, respectively.

3.2.2 Model Specification

We formulate the model in two parts. The first is the marginal distribution of death T_D with a Cox proportional hazards model

$$d\Lambda_{T_D}(t|\mathbf{z}) = \eta(\mathbf{z})dH(t), \tag{3.1}$$

where η characterizes covatiate effects on the terminal event T_D through $\eta(\mathbf{z}) = e^{\beta_\eta \mathbf{z}}$, and the cumulative baseline hazard $H(t)$ summarizes the disease progression pattern leading to death.

The second part of the model is the conditional distribution of incidence T_I given time-to-death T_D. To ensure that incidence always precedes death, integration of the distribution of T_I over $[0, T_D]$ should be 1. If the discrete time survival analysis is considered, it is natural to use a multinomial logit model to characterize this distribution. To deal with continuous time, a continuous counterpart of the multinomial logit model is specified. Assuming time 0 is the reference category, given T_D, the logit compares any time t with the reference category as

$$\log \frac{f_{T_I|T_D}(t|T_D, \mathbf{z})}{Pr(T_I = 0|T_D, \mathbf{z})} = \log \tilde{h}(t) + \beta_\mu(t)\mathbf{z}, \quad 0 < t \leq T_D,$$

where $\beta_\mu(t)$ is a parametric function of time t, representing time-dependent regression coefficients for $T_I = t$; $\tilde{h}(t)$ is an arbitrary function of time t, and $\tilde{h}(t) \geq 0$. Define $\mu(t|\mathbf{z}) = e^{\beta_\mu(t)\mathbf{z}}$, we have

$$\begin{aligned} f_{T_I|T_D}(t|T_D, \mathbf{z}) &= \frac{\tilde{h}(t)\mu(t|\mathbf{z})}{1 + \int_0^{T_D} \tilde{h}(x)\mu(x|\mathbf{z})\, dx}, \quad 0 < t \leq T_D, \\ Pr(T_I = 0|T_D, \mathbf{z}) &= \frac{1}{1 + \int_0^{T_D} \tilde{h}(x)\mu(x|\mathbf{z})\, dx}. \end{aligned} \tag{3.2}$$

For the rest of the chapter, we omit $\mathbf{z}$ as an argument for brevity and denote $\eta(\mathbf{z})$ by

η and $\mu(t|\mathbf{z})$ by $\mu(t)$.

3.2.3 Discretization of the Conditional Incidence Model

For semiparametric estimation, we work with discrete time survival analysis, so it is of interest to discretize the incidence models (3.2).

Given data, incidence events happen at sorted time points

$$0, t_1, t_2, ..., t_n.$$

Split $[0, T_D]$ based on these time points, and obtain intervals as

$$0, (0, t_1], (t_1, t_2], ...$$

and k represents the kth interval. The conditional model can be rewritten as
If $0 < t \leq T_D$,

$$\begin{aligned} Pr(T_I \in [t, t+dt)|T_D) =& \frac{\tilde{h}(t)\mu(t)}{1 + \int_0^{T_D} \tilde{h}(x)\mu(x)\, dx} dt \\ =& \frac{\mu(t)\tilde{h}(t)dt}{1 + \sum\limits_{k:0<t_k\leq T_D} \tilde{h}(t_k)\mu(t_k)d(t_k - t_{k-1})} \end{aligned}$$

Since each interval is infinitely small, $\tilde{h}(t_k)$ and $\mu(t_k)$ is constant within the interval. Define $h(t) \triangleq \tilde{h}(t)dt$, we have

$$\begin{aligned} Pr(T_I \in [t, t+dt)|T_D) &= \frac{h(t)\mu(t)}{1 + \sum\limits_{k:0<t_k\leq T_D} h(t_k)\mu(t_k)}, \quad 0 < t \leq T_D \\ Pr(T_I = t|T_D) &= \frac{1}{1 + \sum\limits_{k:0<t_k\leq T_D} h(t_k)\mu(t_k)}, \quad t = 0. \end{aligned} \tag{3.3}$$

The discrete model takes the form of a multinomial logit model. The summation

of probabilities of T_I falling into any time interval between 0 and T_D is 1. Also, there is a point mass for T_I at time 0, and $h_0\mu_0$ is restricted to 1 for identifiability. Moreover, $\beta_\mu(t)$ is an analog of columns of regression coefficients in the multinomial logit model. However, since the model might be non-identifiable due to the sparse categories, instead of allowing $\beta_\mu(t)$ to be nonparametric at different incidence times, we introduce additional parametric restrictions for $\beta_\mu(t)$ to be a function of time, to ensure the model stability.

3.2.4 Likelihood Construction

We combine (3.1) and (3.3) to build the joint likelihood of cancer diagnosis incidence and death. The likelihood of a single subject with observed data $(X_1, \Delta_1, X_2, \Delta_2)$ falls into one of the following scenarios:

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$):

$$L_{11} = f_{T_D}(X_2) f_{T_I|T_D}(X_1|X_2) = \eta dH(X_2) e^{-\eta H(X_2)} \frac{h(X_1)\mu(X_1)}{1 + \sum\limits_{k:0<t_k\leq X_2} h(t_k)\mu(t_k)}.$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 0$):

   - If $X_2 \geq t_{lf}$,

$$L_{10} = f_{T_D}(\infty) f_{T_I|T_D}(X_1|\infty) = e^{-\eta H(t_{lf2})} \frac{h(X_1)\mu(X_1)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}.$$

- If $X_2 < t_{lf}$,

$$L_{10} = \int_{X_2}^{t_{lf2}} f_{T_D}(t_D) f_{T_I|T_D}(X_1|t_D)\, dt_D + f_{T_D}(\infty) f_{T_I|T_D}(X_1|\infty)$$

$$= \int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{h(X_1)\mu(X_1)}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{h(X_1)\mu(X_1)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}.$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = \Delta_2 = 0$):

- If $X_2 \geq t_{lf1}$,

$$L_{00} = Pr(T_I = \infty, T_D = \infty) = e^{-\eta H(t_{lf2})} \frac{h(\infty)\mu(\infty)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}.$$

- If $X_2 \geq t_{lf2}$ and $X_2 < t_{lf1}$,

$$L_{00} = Pr(T_I > X_2, T_D = \infty) = e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}.$$

- If $X_2 < t_{lf2}$ and $X_2 < t_{lf1}$,

$$L_{00} = Pr(T_I > X_2, X_2 < T_D \leq t_{lf2}) + Pr(T_I > X_2, T_D = \infty)$$

$$= \int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:X_2<t_k\leq t_D} h(t_k)\mu(t_k)}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}.$$

We can alternatively express the joint log-likelihood for a single subject as:

$$\begin{aligned} l &= \Delta_1\Delta_2 \log\left[f_{T_I}(X_1) f_{T_D|T_I}(X_2|X_1)\right] + \Delta_1(1-\Delta_2)\log\left[f_{T_I}(X_1) S_{T_D|T_I}(X_2|T_I = X_1)\right] \\ &\quad + (1-\Delta_1)\log S_{T_I}(X_1) \\ &= \Delta_1 \log f_{T_I}(X_1) + (1-\Delta_1)\log S_{T_I}(X_1) \\ &\quad + \Delta_1\left[\Delta_2 \log f_{T_D|T_I}(X_2|X_1) + (1-\Delta_2)\log S_{T_D|T_I}(X_2|T_I = X_1)\right]. \end{aligned} \tag{3.4}$$

The joint log-likelihood can be partitioned into two parts, such that the contribution from incidence is separated from terminal event. If we can denote l_1 and l_2 as the quantities in each line of equation (3.4), it is easy to see that l_1 is based on information from incidence, while l_2 is based on additional information from the subsequent time segment between incidence and death.

3.2.5 Prediction of Death Given Incidence

The model also allows us to make predictions of the distribution of the time to the terminal event, given observed incidence information. This is of particular interest to clinical practice, as it allows us to predict survival for a subject who has/ has not been diagnosed after some specified time t^*. Derived in Appendix B.1 is the predicted conditional survival functions for death given patient's diagnosis information. Specifically, we have the survival functions

- If $t \geq t^*$,

$$Pr(T_D > t | T_I = t^*) = \frac{\displaystyle\int_t^{t_{lf2}} \frac{\eta e^{-\eta H(t_D)}}{1 + \sum\limits_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + \frac{e^{-\eta H(t_{lf2})}}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}}{\displaystyle\int_{t^*}^{t_{lf2}} \frac{\eta e^{-\eta H(t_D)}}{1 + \sum\limits_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + \frac{e^{-\eta H(t_{lf2})}}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}}$$

for a subject who has been diagnosed at t^*, and

$$Pr(T_D > t | T_I > t^*)$$

$$= \frac{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:t^*<t_k\le t_D} h(t_k)\mu(t_k)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)} dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>t^*} h(t_k)\mu(t_k)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\int_{t^*}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:t^*<t_k\le t_D} h(t_k)\mu(t_k)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)} dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>t^*} h(t_k)\mu(t_k)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}$$

for a subject who has not been diagnosed until t^*.

- If $t < t^*$, since death must happen after incidence, for any subject, whether diagnosed or not at time t^*,

$$Pr(T_D > t | X_1 = t^*, \Delta_1) = 1.$$

3.3 Nonparametric Maximum Likelihood Estimation

The proposed model is semiparametric, consisting of parametric components $\beta = (\beta_\eta, \beta_\mu)$ and nonparametric components $\{h(t)\}$ and $\{H(t)\}$. $h(\cdot)$ is a non-negative function with nonzero values only at the observed incidence times; $H(\cdot)$ is a non-decreasing step function with jumps $\{dH\}$ only at the observed death times. Let us denote the full parameter set $\Omega = (\beta, \{h(t)\}, \{dH(t)\})$.

3.3.1 Martingale Theory

In counting process notation, for subject i, let $N_{1i}(t) = \mathbb{1}(X_{1i} \le t, \Delta_{1i} = 1)$ and $Y_{1i}(t) = \mathbb{1}(X_{1i} \ge t)$ be the observed counting process and at risk process for incidence, respectively; let $N_{2i}(t) = \mathbb{1}(X_{2i} \le t, \Delta_{2i} = 1)$ and $Y_{2i}(t) = \mathbb{1}(X_{2i} \ge t)$ denote the observed counting process and at risk process for death. Log-likelihood

(3.4) can be rewritten in counting process form as

$$l = \sum_i l_i = \sum_i (l_{1i} + l_{2i}), \tag{3.5}$$

where

$$\begin{aligned} l_{1i} &= \int_0^\tau \log d\Lambda_{1i}(t) dN_{1i}(t) - Y_{1i}(t) d\Lambda_{1i}(t), \\ l_{2i} &= \int_0^\tau \Big[\int_{t_1}^\tau \log d\Lambda_{2i}(t|t_1) dN_{2i}(t) - Y_{2i}(t) d\Lambda_{2i}(t|t_1) \Big] dN_{1i}(t_1). \end{aligned}$$

The martingales $dM_{1i}(t)$ and $dM_{2i}(t|t_1)$ can be constructed based on observed counting processes with respect to filtration $\mathcal{F}_i(t-) = \sigma\big\{ N_{1i}(s), N_{2i}(s), Y_{1i}(s), Y_{2i}(s), \boldsymbol{z}_i : s \in [0,t) \big\}$ as

$$\begin{aligned} dM_{1i}(t) &= dN_{1i}(t) - Y_{1i}(t) d\Lambda_{1i}(t), \\ dM_{2i}(t|t_1) &= dN_{2i}(t) dN_{1i}(t_1) - Y_{2i}(t) dN_{1i}(t_1) d\Lambda_{2i}(t|t_1). \end{aligned}$$

Here,

$$\begin{aligned} d\Lambda_{1i}(t) =& \frac{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\mu(t)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)} \, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\mu(t)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\sum_{k:t<t_k\le t_D} h(t_k)\mu(t_k)} h(t) \\ & \frac{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)} \, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\sum_{k:t_k>t} h(t_k)\mu(t_k)} \\ \triangleq& \Theta_{1i}(t) h(t), \\ d\Lambda_{2i}(t|t_1) =& \frac{\eta e^{-\eta H(t)} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:0<t_k\le t} h(t_k)\mu(t_k)}}{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)} \, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}} dH(t) \\ \triangleq& \Theta_{2i}(t; t_1) dH(t). \end{aligned}$$

$d\Lambda_{1i}(t)$ is the hazard of subject i having incidence at time t, and $d\Lambda_{2i}(t|t_1)$ is the

hazard of subject i dying of cancer at time t, given incidence at time t_1. They can be derived through the following probabilistic argument:

$$\mathbb{E}\{dN_{1i}(t)|\mathcal{F}_i(t-)\} = Y_{1i}(t)Pr\{dN_{1i}(t) = 1|Y_{1i}(t) = 1\} = Y_{Ii}(t)d\Lambda_{1i}(t),$$

and $\mathbb{E}\{dN_{2i}(t|t_1)|\mathcal{F}_i(t-)\} = Y_{2i}(t)Pr\{dN_{2i}(t|t_1) = 1|Y_{2i}(t|t_1) = 1\} = Y_{2i}(t)d\Lambda_{2i}(t|t_1)$. (see Appendix B.2 for more details)

3.3.2 Score Functions and NPMLE

Define partial derivatives of $\Theta_{1i}(t)$ and $\Theta_{2i}(t;t_1)$, with respect to $\{h(s)\}$, $\{dH(s)\}$ and β, respectively, as

$$\begin{aligned}
\dot{\Theta}_{1i,h(s)}(t) &= \frac{\partial\Theta_{1i}(t)}{\partial h(s)},\\
\dot{\Theta}_{1i,dH(s)}(t) &= \frac{\partial\Theta_{1i}(t)}{\partial dH(s)},\\
\dot{\Theta}_{1i,\beta}(t) &= \frac{\partial\Theta_{1i}(t)}{\partial \beta},\\
\dot{\Theta}_{2i,h(s)}(t;t_1) &= \frac{\partial\Theta_{2i}(t;t_1)}{\partial h(s)},\\
\dot{\Theta}_{2i,dH(s)}(t;t_1) &= \frac{\partial\Theta_{2i}(t;t_1)}{\partial dH(s)},\\
\dot{\Theta}_{2i,\beta}(t;t_1) &= \frac{\partial\Theta_{2i}(t;t_1)}{\partial \beta}.
\end{aligned}$$

Applying the functional derivative (Hu and Tsodikov, 2013) to the full log-likelihood (3.5), with respect to the infinite-dimensional parameters $\{h(s)\}$ and $\{dH(s)\}$, we can obtain the score functions for $\{h(s)\}$ and $\{dH(s)\}$ as

$$U_{h(s)} = \sum_i \left\{ \frac{dM_{1i}(s)}{h(s)} + \int_s^\tau \frac{\dot{\Theta}_{1i,h(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_s^\tau \int_0^t \frac{\dot{\Theta}_{2i,h(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \right\}, \quad (3.6)$$

$$U_{dH(s)} = \sum_i \left\{ \int_s^\tau \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_s^\tau \int_0^t \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) + \int_0^s \frac{dM_{2i}(s;t_1)}{dH(s)} \right\}, \tag{3.7}$$

which are both martingales under the true model.

Taking derivative of the log-likelihood, with respect to the regression parameter β, we can have the score function for β as

$$U_\beta = \sum_i \left\{ \int_0^\tau \frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_0^\tau \int_0^t \frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \right\}, \tag{3.8}$$

which is also a martingale under the true model.

Set score functions (3.6), (3.7) and (3.8) to be zero, and solve them, can give the NPMLE $\hat{\Omega} = (\hat{\beta}, \{\hat{h}(t)\}, \{d\hat{H}(t)\})$.

3.3.3 Estimation Procedure

It is computationally difficult and costly to directly solve the score equations for the full parameter set Ω to obtain the NPMLE, due to the infinite-dimensional parameters $\{h(t)\}$ and $\{dH(t)\}$ in Ω. Instead, it can be obtained using the profile likelihood approach. This is accomplished by applying an EM algorithm (Tsodikov, 2003) to obtain the implicit estimators for $\{\hat{h}(\beta)\}, \{d\hat{H}(\beta)\}$ that depend on β. To provide computational advantage and increased stability, an artificial latent variable is introduced to transform the multinomial likelihood to Poisson-type (Baker, 1994; Lang, 1996; Tsodikov and Chefo, 2008), and an EM algorithm, treating the artificial variable as missing data, is derived to simplify the maximization step to obtain the maximum likelihood estimators.

We can artificially construct the conditional incidence model as a mixture model

$$f_{T_I|T_D}(t_I|t_D) = E_U\big[f_{T_I|T_D,U}(t_I|t_D,U)\big],$$

where U is a random variable representing artificial missing data, and $U \sim Exp(1)$. Since

$$\begin{aligned} f_{T_I|T_D}(t_I|t_D) =& \frac{h(t_I)\mu(t_I)}{1 + \sum\limits_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)} \\ =& h(t_I)\mu(t_I)E_U\Big\{ \exp\big[-U \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)\big]\Big\}, \end{aligned}$$

we can have

$$f_{T_I|T_D,U}(t_I|t_D,U) = h(t_I)\mu(t_I)\exp\big[-U \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)\big],$$

which takes the form of a Poisson-type likelihood.

After this transformation, an EM algorithm is applied to estimate the nonparametric components $\{dH\}$ and $\{h\}$, holding β fixed. Derivation of the EM algorithm is presented in Appendix B.3. It gives us the score functions for $\{h\}$ and $\{dH\}$ at $(k+1)th$ iteration as

$$U_{h^{(k+1)}}(s) = \frac{dN_1(s)}{h^{(k+1)}(s)} - \Psi_h(s) = 0, \tag{3.9}$$

$$U_{dH^{(k+1)}}(s) = \frac{dN_2(s)}{dH^{(k+1)}(s)} - \Psi_{dH}(s) + \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]\theta_{dH}(s) = 0, \tag{3.10}$$

where

$$
\begin{aligned}
\Psi_{dH}(s) =& \Delta_2 Y_2(s)\eta \\
&+\Delta_1(1-\Delta_2)\mathbb{1}(X_2 < t_{lf2})\Big\{ -\eta + Y_2(s)\eta + [1-Y_2(s)]\frac{\eta[U(s^-, t_{lf2}) + V] - U.s}{U(X_2, t_{lf2}) + V}\Big\} \\
&+(1-\Delta_1)\mathbb{1}(X_1 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\Big\{ -\eta + Y_2(s)\eta \\
&\qquad + [1-Y_2(s)]\frac{\eta[W(s^-, t_{lf2}) + Z] - W.s}{W(X_2, t_{lf2}) + Z}\Big\}, \\
\theta_{dH}(s) =& \mathbb{1}(X_2 < t_{lf2})[1-Y_2(s)](1-\Delta_2)\Big\{\Delta_1\frac{U.s}{U(X_2, t_{lf2}) + V} \\
&\qquad + \mathbb{1}(X_1 < t_{lf1})(1-\Delta_1)\frac{W.s}{W(X_2, t_{lf2}) + Z}\Big\}, \\
\Psi_h(s) =& \Delta_2 Y_2(s)\mu(s)\frac{1}{\sum\limits_{k:0<t_k\le X_2} h(t_k)\mu(t_k)} \\
&+(1-\Delta_1)\mu(s)\Big\{[\mathbb{1}(X_2 \ge t_{lf1}) + \mathbb{1}(t_{lf2} \le X_2 < t_{lf1})Y_2(s)]\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} \\
&\quad + \mathbb{1}(t_{lf2} \le X_2 < t_{lf1})[1-Y_2(s)]\big[\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} - \frac{1}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}\big] \\
&\quad + \mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\frac{-T + Y_2(s)(R+V) - [1-Y_2(s)]Y_{lf2}(s)S_2}{W(X_2, t_{lf2}) + Z}\Big\} \\
&+\Delta_1(1-\Delta_2)\mu(s)\mathbb{1}(X_2 < t_{lf2})\Big\{ -\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} \\
&\quad + \frac{Q + Y_2(s)P(X_2, t_{lf2}) + [1-Y_2(s)]Y_{lf2}(s)P(s^-, t_{lf2})}{U(X_2, t_{lf2}) + V}\Big\}.
\end{aligned}
$$

Here,

$$U.s = \frac{\eta e^{-\eta H(s)}}{1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)},$$

$$U(u,v) = \int_u^v U.s\, dH(s),$$

$$V = \frac{e^{-\eta H(t_{lf2})}}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)},$$

$$W.s = \eta e^{-\eta H(s)} \frac{\sum\limits_{k:X_2<t_k\le s} h(t_k)\mu(t_k)}{1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)},$$

$$W(u,v) = \int_u^v W.s\, dH(s),$$

$$Z = e^{-\eta H(t_{lf2})} \frac{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)},$$

$$P.s = \frac{\eta e^{-\eta H(s)}}{[1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)]^2} = \frac{U.s}{1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)},$$

$$P(u,v) = \int_u^v P.s\, dH(s),$$

$$Q = \frac{e^{-\eta H(t_{lf2})}}{[1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)]^2} = \frac{V}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)},$$

$$R.s = \eta e^{-\eta H(s)} \frac{\sum\limits_{k:X_2<t_k\le s} h(t_k)\mu(t_k)}{[1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)]^2} = \frac{W.s}{1 + \sum\limits_{k:0<t_k\le s} h(t_k)\mu(t_k)},$$

$$R = \int_{X_2}^{t_{lf2}} R.s\, dH(s),$$

$$S_2 = [1 + \sum_{k:0<t_k\le X_2} h(t_k)\mu(t_k)] \int_{s^-}^{t_{lf2}} P.s\, dH(s),$$

$$T = \frac{[1 + \sum\limits_{k:0<t_k\le X_2} h(t_k)\mu(t_k)]e^{-\eta H(t_{lf2})}}{[1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)]^2} = [1 + \sum_{k:0<t_k\le X_2} h(t_k)\mu(t_k)]Q.$$

Solving equations (3.9) and (3.10), we can have Breslow-type estimators

$$h^{(k+1)}(s) = \frac{\sum_i dN_{1i}(s)}{\sum_i \Psi_{h,i}^{(k)}(s)}, \tag{3.11}$$

$$dH^{(k+1)}(s) = \frac{\sum_i dN_{2i}(s) + \Big[\sum_i \theta_{dH,i}^{(k)}(s)\Big] dH^{(k)}(s)}{\sum_i [\Psi_{dH,i}^{(k)}(s) + \theta_{dH,i}^{(k)}(s)]}. \tag{3.12}$$

Equations (3.11) and (3.12) solve for $h(s)$ and $dH(s)$ iteratively, $k = 0, 1, 2, ...$, until convergence, i.e., $h^{(k+1)} \to h^{(k)}$, $dH^{(k+1)} \to dH^{(k)}$. Note, at convergence, the last term in equations (3.9) and (3.10) disappears, and the estimating equations are the same as that obtained from observed data. Estimators at convergence are consistent (Tsodikov (2003)).

The estimation procedure is described as follows:

Start with $\beta^{(0)} = 0$, $j = 0$.

1. Maximize the likelihood over $h(\beta)$ and $H(\beta)$, respectively, given $\beta = \beta^{(j)}$:

   (a) Set $k = 0$. Initialize $\hat{h}^{(0)}(s)$ and $d\hat{H}^{(0)}(s)$ with any positive value, e.g. 1, at observed incidence and death times, respectively.

   (b) With β fixed, calculate $\hat{h}^{(k+1)}(s)$ and $d\hat{H}^{(k+1)}(s)$ using equations (3.11) and (3.12).

   (c) Repeat step (b) to update $\hat{h}^{(k+1)}(s)$ and $d\hat{H}^{(k+1)}(s)$, until convergence

$$\left\|\hat{h}^{(k+1)}(s) - \hat{h}^{(k)}(s)\right\|_2 < \epsilon \quad \text{and} \quad \left\|d\hat{H}^{(k+1)}(s) - d\hat{H}^{(k)}(s)\right\|_2 < \epsilon.$$

2. Maximize the profile log-likelihood $l_{pr}(\beta) = l(\beta, \{\hat{h}(\beta)\}, \{d\hat{H}(\beta)\})$ over β:

   (a) Calculate profile log-likelihood $l(\beta, \{\hat{h}(\beta)\}, \{d\hat{H}(\beta)\})$.

   (b) Find $\beta^{(j+1)}$ by maximizing $l_{pr}(\beta)$ over β: $\beta^{(j+1)} = \text{argmax}_\beta\, l_{pr}(\beta)$, us-

ing conventional optimization method, e.g. Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS).

Iteratively apply steps 1-2 to estimate β, until convergence of $l_{pr}(\beta)$

$$l_{pr}(\beta^{(j+1)}) - l_{pr}(\beta^{(j)}) < \xi$$

Note the convergence tolerance for the inner loop (step 1) should be stricter than that for the outer loop (step 2), e.g. $\epsilon = 10^{-6}$, $\xi = 10^{-5}$.

3.4 Asymptotic Properties

We apply the empirical process (Kosorok, 2008; Van Der Vaart and Wellner, 2000) and the theory of martingale structure in counting process to build the asymptotic properties, adapted from previous work (Zeng and Lin, 2007, 2010; Chen, 2009, 2010; Hu and Tsodikov, 2014; Rice and Tsodikov, 2017).

Assuming regularity conditions hold, in the following, Theorem III.1 and Theorem III.2 state the consistency and weak convergence results of the NPMLE $\hat{\Omega} = (\hat{\beta}, \{\hat{h}\}, \{d\hat{H}\})$, while Theorem III.3 justifies the use of negative Hessian matrix from profile log-likelihood in variance estimation. Regularity conditions and proofs are similar as those in Chap II.

Under regularity conditions,

Theorem III.1. *With probability 1: $\hat{\beta}$ converges to β^0; $\hat{h}(t)$ and $\hat{H}(t)$ converges to $h^0(t)$ and $H^0(t)$ uniformly over the interval $[0, \tau]$, respectively. Here, β^0, $h^0(t)$ and $H^0(t)$ are the true values of $\hat{\beta}$, $\hat{h}(t)$ and $\hat{H}(t)$.*

Theorem III.2. *$n^{1/2}\{\hat{\beta} - \beta^0, \hat{h}(t) - h^0(t), \hat{H}(t) - H^0(t)\}$ converges weakly to a zero-*

mean Gaussian process. In addition, consider a linear functional of $\hat{\Omega}$,

$$n^{1/2}\Big\{a^T(\hat{\beta}-\beta^0)+\int_0^\tau \Big[b_1(t)d\big(\hat{h}(t)-h^0(t)\big)+b_2(t)d\big(\hat{H}(t)-H^0(t)\big)\Big]\Big\},$$

where a *is a real vector,* $b_1(t)$ *and* $b_2(t)$ *are functions with bounded total variation in* $[0,\tau]$*, evaluated at the observed incidence and death times, respectively. Let* $\Phi^T = (a^T, \{b_1(\cdot)\}^T, \{b_2(\cdot)\}^T)$*. The asymptotic variance-covariance function of the linear functional above can be consistently estimated by* $\Phi^T(\mathcal{I}_n)^{-1}\Phi$*, where* $\mathcal{I}_n$ *is the observed information matrix for* Ω*, that is,* $\mathcal{I}_n = -\dfrac{\partial^2 l_n}{\partial\Omega\partial\Omega^T}\Big|_{\Omega=\hat{\Omega}}$*, where* $l_n = n^{-1}\sum_{i=1}^n l_i$*,* l_i *is subject* i*'s observed log-likelihood defined as equation (3.5).*

Theorem III.3. *The inverse of the negative Hessian matrix of the profile log-likelihood with respect to* β *is a consistent estimator of the limiting variance-covariance matrix of* $\hat{\beta}$*. That is,*

$$\Big(-\frac{\partial^2 l_{pr,n}}{\partial\beta\partial\beta^T}\Big|_{\beta=\hat{\beta}}\Big)^{-1} \xrightarrow{p} Var\big[\sqrt{n}(\hat{\beta}-\beta^0)\big],$$

where $l_{pr,n} = n^{-1}\sum_{i=1}^n l_{pr,i}(\beta) = n^{-1}\sum_{i=1}^n l_i\big(\beta, \{\hat{h}(\beta)\}, \{d\hat{H}(\beta)\}\big)$.

3.5 Simulation Studies

This section presents Monte Carlo simulations to assess our proposed methodology. The simulation settings were as follows. The true $h(t) = 30t$, and the true cumulative baseline hazard for terminal event was $H(t) = 0.1t^{2.5}$. We considered two covariates Z_1 and Z_2, where $Z_1 \sim Bernoulli(0.5)$, and $Z_2 \sim Normal(0,1)$. Covariates entered the model via $\eta(\mathbf{z}) = e^{\beta_1 z_1 + \beta_2 z_2}$, and $\mu(t|\mathbf{z}) = e^{(\beta_3 t + \beta_4)z_1}$, with true parameters $(\beta_1, \beta_2, \beta_3, \beta_4) = (0.5, -1, 0.8, -1)$. Censoring was simulated from the exponential distribution $Exp(0.6)$, yielding 50% intermediate and 65% terminal events censored.

3.5.1 Finite-sample Properties of Parameter Estimates

We conducted simulations to examine the finite-sample properties of the parameter estimates obtained. Samples of size 300 and 500 were examined, each with 1000 replicates. Standard errors were obtained from the numerically evaluated Hessian matrix at the solution.

The simulation results are summarized in Table 3.1. The proposed estimators are almost unbiased and get more accurate as sample size increases. The asymptotic standard errors (ASE) are close to the empirical standard deviations (ESD), suggesting reasonable approximation of the variance estimators; and with larger sample size, standard errors are smaller, and we see better agreement between ASE and ESD. The 95% coverage probabilities for all the estimators approach the 95% nominal level for both sample sizes.

Table 3.1: Simulation results using the proposed semiparametric joint model.

N	β	Truth	Bias	ASE	ESD	95% CP
300	β_1	0.5	-0.012	0.191	0.197	0.963
	β_2	-1	0.023	0.111	0.116	0.948
	β_3	0.8	0.084	0.445	0.406	0.942
	β_4	-1	-0.168	0.767	0.716	0.950
500	β_1	0.5	-0.001	0.147	0.151	0.961
	β_2	-1	0.013	0.087	0.089	0.956
	β_3	0.8	-0.014	0.315	0.299	0.930
	β_4	-1	-0.030	0.548	0.524	0.937

ASE: average of estimated standard errors
ESD: empirical standard deviation based on Monte Carlo estimates
95% CP: 95% coverage probability

3.5.2 Hypothesis Testing

Given the proposed estimators, we constructed the likelihood ratio test (LRT) to test the null hypothesis:

H_0: z_1 does not have an effect on the marginal terminal event.

The simulation set-up was the same as before. Sample size was chosen to be 500. The simulation procedure was as follows:

1. Generate a data set under H_0.
2. Fit the models with and without covariate z_1 entering η, respectively, and obtain the two log-likelihoods.
3. Calculate the LRT statistic, and decide if H_0 is rejected.
4. Repeat steps 1-3 1000 times, and calculate the empirical p-value.

Based on the simulations, the 0.05 significance level is well maintained (0.048).

3.6 Application to Prostate Cancer Screening Trial Data

3.6.1 Data

The proposed method was applied to the motivating setting of the prostate cancer screening trial, testing the screening effect on cancer mortality. The data come from the PLCO trial, with patients entering the trial aged 55-74 years old. The control arm in the PLCO trial was contaminated (Vickers, 2017), as about 50% of PLCO control patients had PSA testing before enrollment, and of the remainder, close to 90% had PSA measured during the trial (Shoag et al., 2016). Thus, we needed to introduce a set of uncontaminated control data from external data to assess the screening effect. Comparing SEER with PLCO data (Pinsky et al., 2012), there was no unambiguous evidence showing a healthy volunteer effect. Therefore, a simulated subset of SEER data, with diagnosis between 1980 and 1987 before the use of PSA as a screening tool, was created to act as uncontaminated "perfect" controls.

In the combined data set, 76,674 subjects are from PLCO trial (38,335 subjects in the screening arm, and 38,339 subjects in the control arm), and 38,335 subjects are from the SEER control arm. For patients from the PLCO trial, 4418 (11.52%) in

the screening arm were diagnosed with prostate cancer, and 145 (0.38%) died of it; while 4036 (10.53%) in the control arm were diagnosed and 142 (0.37%) died of it. For patients in the simulated SEER control arm, 2726 (7.11%) were diagnosed with prostate cancer, and 606 (1.58%) died of it. Maximum follow-up time was 13 years. From the data, we have information about time-to-diagnosis and time-to-death. In addition to arm (screening/control) and trial (PLCO/SEER), we are also interested in studying the age effect, since age is an important risk factor for prostate cancer. To make patients from different trials comparable, the simulated SEER data have the same age distribution as the PLCO data.

3.6.2 Results

In our proposed method, covariates entered the model through

$$\eta(t|\mathbf{z}) = e^{\beta_{\eta 1} Arm + \beta_{\eta 2} Trial + \beta_{\eta 3} Age}$$

$$\mu(t|\mathbf{z}) = e^{(\beta_{\mu 1} + \beta_{\mu 2} t) Arm + (\beta_{\mu 3} + \beta_{\mu 4} t) Trial + (\beta_{\mu 5} + \beta_{\mu 6} t) Age}$$

The goodness of fit for the Cox regression model for marginal death was checked. The proportional hazards assumptions for the covariates age, arm and trial are supported, with respective goodness of fit p-values of 0.72, 0.92 and 0.58, respectively. Table 3.2 presents the regression coefficient estimates for both the proposed joint model and the Cox model for marginal death. Based on the estimates of our proposed joint model, older patients have greater risk of death (β=0.55, HR=1.73, p-value<0.0001), and screening helps reduce cancer mortality (β=-1.30, HR=0.27, p-value<0.0001). Due to the relatively small number of deaths, which may lead to lack of power to detect any difference between arms, and the contamination of PLCO control arm, there is no significant difference between PLCO screening and PLCO control arms.

Regarding the analysis of time-to-death, theoretically, point estimates of β_η's

should be similar for the two models; as the table shows, the model results are as expected. The standard errors for β_η's in the proposed method are slightly smaller than that from the marginal model, and this efficiency gain comes from information of subjects with incidence observed, yet death censored.

Table 3.2: Analysis of prostate cancer screening trial data.

(a) Log-hazard ratio for time-to-death (β_η)

	Proposed Joint Model			Cox for Marginal Death		
Parameter	Est.	SE	p-value	Est.	SE	p-value
Age	0.547	0.032	< 0.0001	0.551	0.032	< 0.0001
PLCO scr vs. PLCO control	0.039	0.117	0.738	0.035	0.118	0.767
PLCO control vs. SEER control	-1.303	0.092	< 0.0001	-1.274	0.093	< 0.0001

(b) Parameter estimate for time-to-incidence given death (β_μ)

	Proposed Joint Model		
Parameter	Est.	SE	p-value
$\beta_{\mu 1}$	-0.003	0.001	0.002
$\beta_{\mu 2}$	-1.174	0.102	< 0.0001
$\beta_{\mu 3}$	-0.056	0.002	< 0.0001
$\beta_{\mu 4}$	-0.084	0.153	0.584
$\beta_{\mu 5}$	-0.013	0.001	< 0.0001
$\beta_{\mu 6}$	-0.085	0.040	0.036

To assess model fit (Dejardin et al., 2010), Figure 3.1 and 3.2 present the survival estimates for time-to-diagnosis and time-to-death, respectively, for subjects who entered the study at age 68, stratified by arm and trial. Both of the figures have plots of marginal survival functions obtained using the proposed model, which closely match the KM estimates obtained from the observed data, confirming the validity of the model.

3.6.3 Prediction of Death Given Diagnosis

The proposed joint model can predict cancer mortality given diagnosis information, which is of particular interest to clinical practice. As shown in Section 3.2.5, it allows us to calculate this conditional survival function and produce the plots shown in Figure 3.3. Figure 3.3 presents the predicted survival rates stratified by arm and

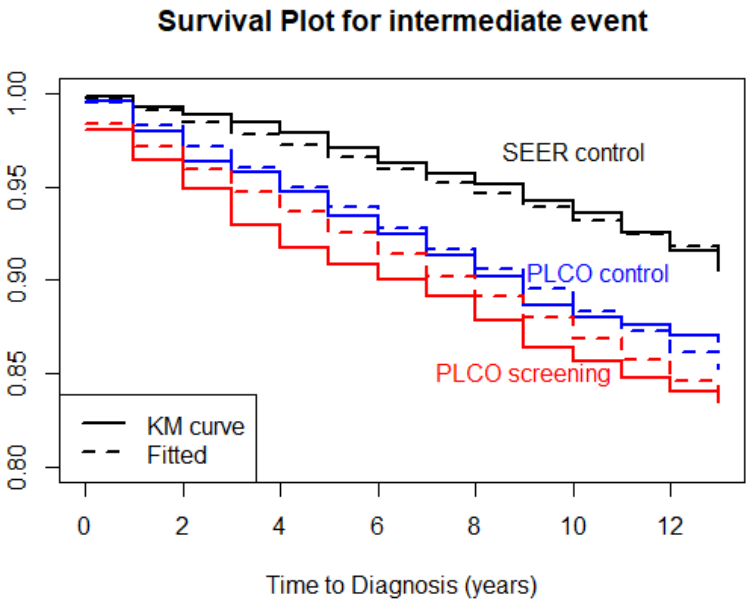

Figure 3.1: Survival estimates for time-to-diagnosis for subjects who entered the study at age 68. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

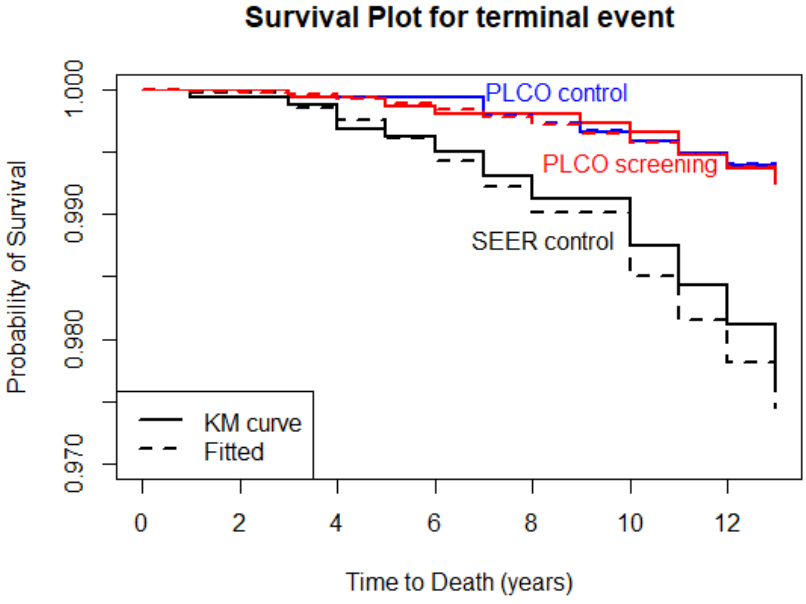

Figure 3.2: Survival estimates for time-to-death for subjects who entered the study at age 68. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

trial, given diagnosis status (diagnosed or not diagnosed) after 1 year of randomization, among men who entered the study at age 68. We can see that given the same diagnosis information, patients without screening (SEER control) tend to die of prostate cancer earlier, compared with those with screening (PLCO trial). The PLCO screening arm and PLCO control arm present no surprise and show little difference in predicted survival rates. In addition, note that the survival probabilities for patients with NO cancer diagnosed at 1 year (right plot) are very close to 1, while those with cancer diagnosed (left plot) are not. Thus, in comparison with patients with NO cancer detected, patients with cancer detected are more likely to die from it, as expected.

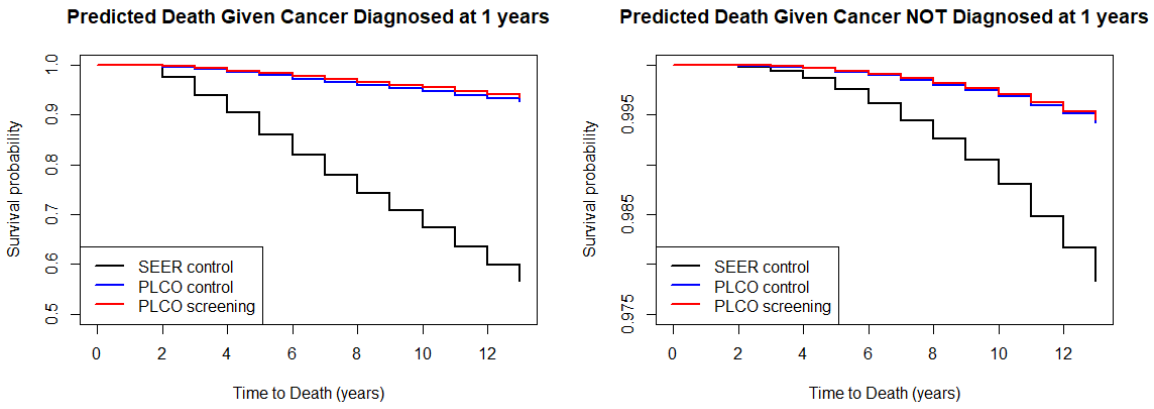

Figure 3.3: Predicted survival estimates for death from prostate cancer, given diagnosis status after 1 year of randomization, among men who entered the study at age 68. Left figure is for a hypothetical subject who has prostate cancer diagnosed at the time indicated, while the right one is for a subject not diagnosed at that time.

3.7 Discussion

We have presented a framework to test the screening effect on cancer mortality with a semiparametric joint model of ordered events. Within the proposed model, the NPMLE can be obtained by EM algorithm and profile likelihood, and its asymptotic properties established. Simulation studies indicate good finite sample performance.

Application of the proposed model to the PLCO prostate cancer data, combined with simulated SEER control data, reveals the benefit of screening on reducing cancer mortality.

The model framework is similar to that of Chapter II. It separately specifies marginal death and conditional incidence given death models. The marginal death distribution allows us to test the screening effect on the terminal event in an intuitive way. Regarding the conditional incidence distribution, we extend the beta model in Chapter II to a semiparametric model based on a multinomial logit model. The technical difficulty comes from the complexity of the multinomial likelihood function. EM algorithm can be viewed as a way to replace maximization of the original observed likelihood by the corresponding complete-data likelihood, which offers computational advantage and increased stability. With this idea in mind, an artificial latent variable is introduced to transform the multinomial likelihood to Poisson-type by means of expectation, and an EM algorithm, treating the artificial variable and censored observations as missing data, is derived to simplify the maximization step to obtain the maximum likelihood estimators. Compared with the beta conditional model in Chapter II, this semiparametric model relaxes parametric assumptions and thus provides wider applicability and increased robustness.

Compared with the model on marginal death, this joint model provides a better understanding of the natural history of a disease. This carries clinical significance since the joint model can be used to predict disease prognosis and design optimal treatments. In addition, it has slight efficiency gain, due to the extra information from observations with incidence observed, yet death censored. In future work, it would be of interest to further improve efficiency, by utilizing the strong relationship between the two events driven by a common cancer progression process.

CHAPTER IV

A Mechanistic Joint Model to Investigate the Screening Effect on Cancer Mortality

4.1 Introduction

Prostate-specific-antigen (PSA) testing, a screening tool for prostate cancer, has been widely used in the United States since the late 1980s. Treatment for prostate cancer can be confounded by indication provided by the cancer diagnosis that can occur earlier with PSA screening. While PSA tests led to increased incidence of cancer diagnosis, their effect on prostate cancer mortality is disputed (Andriole et al., 2012).

In its natural history, a disease usually progresses through multiple events over time. Specifically, for cancer studies, there are many events of interest, such as time-to-diagnosis, time-to-remission, time-to-recurrence, and time-to-death. A joint model approach is thus needed to capture the relationship between the events driven by a common cancer progression process.

We focus on two sequential events, time-to-diagnosis and time-to-death. In the previous two chapters, we evaluated whether screening can benefit cancer mortality, despite an undoubted effect on cancer diagnosis. In this chapter, we aim to explore a methodology to provide efficiency gains to test the screening effect. It is critical to

establish the relationship between cancer diagnosis and death.

Copula modeling has become a popular tool in multivariate analysis of correlated data since the fundamental work of Clayton (1978). Oakes (1994) and Genest et al. (1995) proposed the semiparametric two-stage estimation approach. Hsieh et al. (2008) and Chen (2012) suggested using semiparametric transformation models for the marginal distributions and a copula model to link them for the joint distribution. However, due to the dependence between intermediate and terminal events induced by cancer diagnosis always preceding death, we cannot make the covariate affect one event while not affecting the other, which excludes the null hypothesis.

Another strategy to formulate the dependence between the events is the conditional specification based on modulated point process (Cook and Lawless, 2007; Hu and Tsodikov, 2013; Rice and Tsodikov, 2017). This strategy describes how the occurrence of a latent event affects the risk of occurrence of future events in a time-dependent fashion with a stochastic process frailty. Unlike frailty models which can only handle positive correlations (Hougaard, 2012), it can accommodate both positive and negative associations. In addition, this frailty is no longer a random variable, but rather a stochastic process that jumps from 0 to 1 at the time of the latent event, to ensure that the subject is not at risk of the future event until the occurrence of the latent event. This formulation can thus guarantee the ordering of the events.

Figure 4.1 presents the disease progression process we consider, where both incidence of cancer diagnosis and death are observed, yet the type of incidence (causal to death or not) is missing. Overdiagnosis is caused if non-causal incidence is detected. This mechanism thus incorporates the null hypothesis, where screening does not affect cancer mortality, although it increases the risk of the incidence event. A mechanistic joint model is formulated on the partially observed disease progression process, and a test is developed for the causal effect of screening. We capture the dependence between events by using a jump process as the frailty. This formula-

tion results in a strong link and information sharing between cancer incidence and death, thus providing efficiency gains. Statistical inference is based on nonparametric maximum likelihood estimation with EM algorithm and profile likelihood.

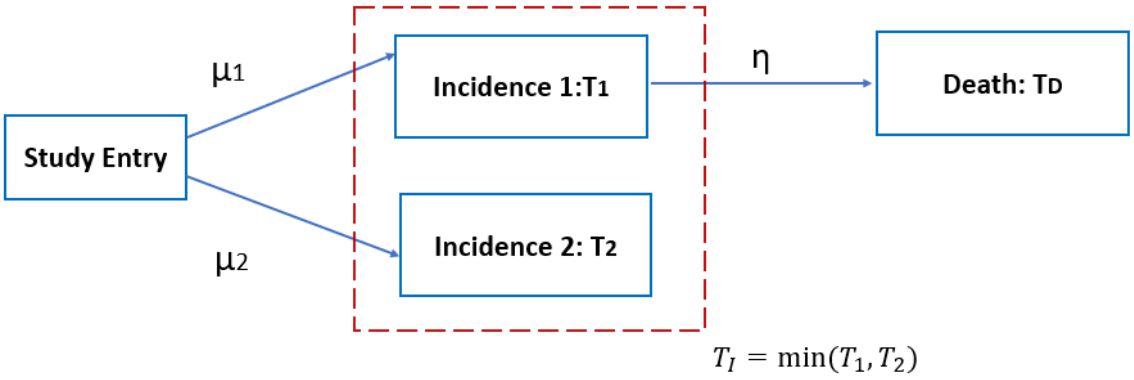

Figure 4.1: Diagram of cancer progression: Patient is diagnosed with cancer at time T_I and dies at T_D, yet the incidence type (incidence 1 or 2) is always missing.

The rest of the chapter is organized as follows. Section 4.2 describes the framework of our proposed model and derives its essential distributional characteristics. Section 4.3 presents the likelihood in counting process form and the corresponding martingale properties, as well as the estimating procedure for the nonparametric maximum likelihood estimators. The asymptotic properties are outlined in Section 4.4, with proofs given in the Appendix C.6. Section 4.5 conducts simulation studies, evaluating the performance of the proposed estimators and hypothesis testing with finite samples. Section 4.6 analyzes the prostate cancer data. Finally, we discuss the results in Section 4.7.

4.2 Statistical Framework

4.2.1 Data Structure and Notation

Consider sequential events, such as cancer incidence and death, in our model. Let T_I and T_D be the time to incidence and death, respectively; $\mathbf{z}$ be a set of fully observed

covariates; and C be the censoring time which is independent of T_I and T_D given $\mathbf{z}$. T_I results from two competing incidences: incidence 1 and incidence 2, whose times to event are T_1 and T_2, respectively, and are not directly observed. Suppose type of incidence is never observed. By definition, $T_I = \min(T_1, T_2)$. We also assume that of the two incidences, incidence 1 is causal incidence, while incidence 2 is non-causal, i.e. only incidence 1 may cause death. Based on this assumption, the unobserved causal incidence must precede the terminal event: $T_D \geq T_1 \ \ w.p.\ 1$. In addition, we assume $Pr(T_1 = T_D) = 0$.

We observe $(X_I, \Delta_I, X_D, \Delta_D)$, where $X_I = \min(T_I, C)$ is the time to the intermediate event (i.e. incidence or censoring); $X_D = \min(T_D, C)$ is the time to the terminal event (i.e. death or censoring); $\Delta_I = \mathbb{1}(X_I = T_I)$ is the indicator of observing incidence; and $\Delta_D = \mathbb{1}(X_D = T_D)$ is the indicator of observing death.

4.2.2 Model Specification

We formulate our model in two parts. The first is the marginal hazard of incidence:

$$d\Lambda_I(t|\mathbf{z}, Tx) = \lim_{h \to 0} \frac{P(T_I \in [t, t+h) | T_I \geq t, \mathbf{z}, Tx)}{h} = \mu dH_I(t),$$

here, Tx is the covariate of screening, $\mathbf{z}$ represents covariates of interest other than screening, and they enter the marginal incidence model through $\mu = \exp(\beta_{Tx} Tx + \beta_{\mathbf{z}} \mathbf{z})$.

Hazard of incidence can be partitioned into two parts, hazard of causal (incidence 1) and non-causal (incidence 2) incidence, $d\Lambda_1$ and $d\Lambda_2$, respectively. Suppose the partition only depends on Tx, for example, $\dfrac{d\Lambda_1}{d\Lambda} = e^{-\gamma Tx}$. Note that γ should be between 0 and 1 to guarantee that the proportion of causal incidence falls between 0

and 1. Then,

$$
\begin{aligned}
d\Lambda_1(t|\mathbf{z}, Tx) =& \lim_{h\to 0} \frac{P(T_1 \in [t, t+h)|T_1 \geq t, \mathbf{z}, Tx)}{h} \\
=& \exp\left[(\beta_{Tx} - \gamma)Tx + \beta_{\mathbf{z}}\mathbf{z}\right] dH_I(t) = \mu_1 dH_I(t), \qquad (4.1) \\
d\Lambda_2(t|\mathbf{z}, Tx) =& \lim_{h\to 0} \frac{P(T_2 \in [t, t+h)|T_2 \geq t, \mathbf{z}, Tx)}{h} \\
=& (1 - e^{-\gamma Tx}) \exp\left(\beta_{Tx}Tx + \beta_{\mathbf{z}}\mathbf{z}\right) dH_I(t) = \mu_2 dH_I(t). \qquad (4.2)
\end{aligned}
$$

The second part of the model is the conditional hazard of death given time to causal incidence $d\Lambda_D$:

$$
d\Lambda_D(t|T_1, \mathbf{z}) = \lim_{h\to 0} \frac{P(T_D \in [t, t+h)|T_D \geq t, T_1, \mathbf{z})}{h} = \mathbb{1}(t \geq T_1)\eta dH_D(t), \qquad (4.3)
$$

where $\eta = \exp\left(\beta^D_{Tx}Tx + \beta^D_{\mathbf{z}}\mathbf{z}\right)$.

To account for potentially different patterns of disease progression for the time to incidence and the time to death, separate nonparametrically cumulative baseline hazard functions H_I, H_D are used. The cumulative baseline hazard $H_I(t)$ summarizes the temporal disease progression pattern leading to an incidence. The other cumulative baseline hazard $H_D(t)$ summarizes the process leading to death. μ_1 models covariate effects on the time to causal incidence T_1, μ_2 models covariate effects on the time to non-causal incidence T_2, and η models covariate effects on the failure time T_D. Covariate of screening Tx and other covariates of interest $\mathbf{z}$ enter the model through μ_1, μ_2 and η, and $\boldsymbol{\beta} = (\beta_{\mathbf{z}}, \beta_{Tx}, \gamma, \beta^D_{\mathbf{z}}, \beta^D_{Tx})'$ is the combined vector of regression coefficients.

The proposed model belongs to a class of stochastic process frailty models, with unobserved stochastic process $\mathbb{1}(t \geq T_1)$ in (4.3) acting multiplicatively on the baseline hazard of death. It can also be used to test the screening effect. Specifically, the null

hypothesis of no benefit of screening on cancer mortality is:

$$H_0 : \beta_{Tx} = \gamma, \beta_{Tx}^D = 0.$$

4.2.3 Conditional Likelihood

Based on the proposed model hazard functions (4.1), (4.2) and (4.3), the conditional likelihood L_0 of observed data, given unobserved time to causal incidence T_1, can be derived as:

1. Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$L_0 = \begin{cases} \eta dH_D(X_D) e^{-\mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(X_I)]}, & T_1 = X_I \\ \eta\mu_2 dH_I(X_I) dH_D(X_D) e^{-\mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(t_1)]}, & T_1 \in (X_I, X_D] \\ 0, & otherwise. \end{cases} \tag{4.4}$$

2. Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$L_0 = \begin{cases} 0, & T_1 < X_I \\ e^{-\mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(X_I)]}, & T_1 = X_I \\ \mu_2 dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(t_1)]}, & T_1 \in (X_I, X_D] \\ \mu_2 dH_I(X_I) e^{-\mu_2 H_I(X_I)}, & T_1 \in (X_D, \infty). \end{cases} \tag{4.5}$$

3. Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$L_0 = \begin{cases} 0, & T_1 \leq X_I \\ e^{-\mu_2 H_I(X_I)}, & T_1 > X_I. \end{cases} \tag{4.6}$$

4.2.4 Likelihood Construction

The time to causal incidence T_1 is unobserved, and its density function is

$$f_1(t_1) = \mu_1 dH_I(t_1) e^{-\mu_1 H_I(t_1)}$$

The likelihood for a single subject with observed data $(X_I, \Delta_I, X_D, \Delta_D)$ can be obtained by taking expectation of the conditional likelihood L_0 over T_1. The contribution of each subject to the likelihood falls into one of the following scenarios:

1. Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$\begin{aligned} L_{11} &= \int_{X_I}^{X_D} f_1(t_1) f_2(X_I) f_D(X_D|t_1) dt_1 + f_1(X_I) S_2(X_I) f_D(X_D|X_I) \\ &= \eta\mu_1 dH_I(X_I) dH_D(X_D) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} \Bigg[\mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) \\ &\qquad + e^{\eta H_D(X_I) - \mu_1 H_I(X_I)} \Bigg] \end{aligned} \tag{4.7}$$

2. Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$\begin{aligned} L_{10} &= f_1(X_I) S_2(X_I) S_D(X_D|X_I) + S_1(X_D) f_2(X_I) + \int_{X_I}^{X_D} f_1(t_1) f_2(X_I) S_D(X_D|t_1) dt_1 \\ &= dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} \Bigg[\mu_1 e^{\eta H_D(X_I) - \mu_1 H_I(X_I)} + \mu_2 e^{\eta H_D(X_D) - \mu_1 H_I(X_D)} \\ &\qquad + \mu_1\mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) \Bigg] \end{aligned} \tag{4.8}$$

3. Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$L_{00} = S_1(X_I) S_2(X_I) = e^{-(\mu_1 + \mu_2) H_I(X_I)} \tag{4.9}$$

Combining (4.7), (4.8) and (4.9), the contribution of subject i to the observed data likelihood is:

$$L_i = \Delta_I \Delta_D L_{11,i} + \Delta_I (1 - \Delta_D) L_{10,i} + (1 - \Delta_I) L_{00,i} \tag{4.10}$$

4.2.5 Conditional Distribution of Time to Causal Incidence

The model also allows us to make predictions of the distribution of time to causal incidence, given observed data and estimates of $\eta, \mu_1, \mu_2, \{dH_I\}$ and $\{dH_D\}$. This is of particular interest to clinical practice, as it allows us to examine, for a subject who has not experienced the terminal event after some specified time, the distribution of the time to the causal incidence. Specifically, the survival functions can be written as (see Appendix C.2 for details):

1. Consider a subject whose incidence and death are both observed,

$$S_{T_1}(t_1|X_I, X_D) = \begin{cases} 1, & t_1 < X_I \\ \dfrac{\mu_2 \int_{t_1}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x)}{e^{\eta H_D(X_I) - \mu_1 H_I(X_I)} + \mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x)}, & t_1 \in [X_I, X_D] \\ 0, & t_1 > X_D. \end{cases}$$

2. Consider a subject whose incidence is observed, yet terminal event censored,

   - If $t_1 < X_I$,

   $$S_{T_1}(t_1|X_I, X_D) = 1.$$

   - If $t_1 \in [X_I, X_D]$,

$$\begin{aligned} & S_{T_1}(t_1|X_I, X_D) \\ = & \frac{\mu_1 \mu_2 \int_{t_1}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) + \mu_2 e^{\eta H_D(X_D) - \mu_1 H_I(X_D)}}{\mu_1 e^{\eta H_D(X_I) - \mu_1 H_I(X_I)} + \mu_1 \mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) + \mu_2 e^{\eta H_D(X_D) - \mu_1 H_I(X_D)}}. \end{aligned}$$

- If $t_1 > X_D$,

$$S_{T_1}(t_1|X_I, X_D)$$
$$= \frac{\mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(t_1)}}{\mu_1 e^{\eta H_D(X_I)-\mu_1 H_I(X_I)} + \mu_1\mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x)-\mu_1 H_I(x)} dH_I(x) + \mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(X_D)}}.$$

3. Consider a subject with no events observed,

$$S_{T_1}(t_1|X_I, X_D) = \begin{cases} 1, & t_1 < X_I \\ e^{-\mu_1[H_I(t_1)-H_I(X_I)]}, & t_1 \geq X_I. \end{cases}$$

4.3 Nonparametric Maximum Likelihood Estimation

The proposed model is semiparametric, consisting of a parametric component $\boldsymbol{\beta}$ for covariate effects and nonparametric components $H_I(\cdot), H_D(\cdot)$ for baseline hazards. H_I and H_D are non-decreasing step functions with jumps $\{dH_I\}$ and $\{dH_D\}$ only at the observed event (incidence/death) times. Let us denote the full parameter set $\Omega = (\boldsymbol{\beta}, \{dH_I\}, \{dH_D\})$. We use the EM algorithm (Tsodikov (2003)) and profile likelihood approach to obtain the nonparametric maximum likelihood estimator (NPMLE) for Ω.

4.3.1 Martingale Theory

In counting process notation, for subject i, let $N_{Ii}(t) = \mathbb{1}(X_{Ii} \leq t, \Delta_{Ii} = 1)$ and $Y_{Ii}(t) = \mathbb{1}(X_{Ii} \geq t)$ be the observed counting process and at risk process for incidence, respectively; let $N_{Di}(t) = \mathbb{1}(X_{Di} \leq t, \Delta_{Di} = 1)$ and $Y_{Di}(t) = \mathbb{1}(X_{Di} \geq t)$ denote the observed counting process and at risk process for death.

We can define the following martingales, based on observed counting processes

with respect to filtration $\mathcal{F}_i(t-) = \sigma\big\{N_{Ii}(x), N_{Di}(x), Y_{Ii}(x), Y_{Di}(x), \boldsymbol{z}_i : x \in [0,t)\big\}$,

$$\begin{aligned} dM_{Ii}(t) &= dN_{Ii}(t) - Y_{Ii}(t)d\Lambda_{Ii}(t) \\ &= dN_{Ii}(t) - Y_{Ii}(t)\Theta_{Ii}\big(t;\beta,\overline{H_I}(t)\big)dH_I(t), \\ dM_{Di}(t) &= dN_{Di}(t) - Y_{Di}(t)d\Lambda_{Di}(t;t_I) \\ &= dN_{Di}(t) - Y_{Di}(t)\Theta_{Di}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big)dH_D(t), \end{aligned}$$

where

$$\begin{aligned} &\Theta_{Ii}\big(t;\beta,\overline{H_I}(t)\big) = \mu_1 + \mu_2, \\ &\Theta_{Di}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big) \\ =&\eta\mu_1 \frac{e^{\eta H_D(t_I)-\mu_1 H_I(t_I)} + \mu_2\int_{t_I}^{t} e^{\eta H_D(x)-\mu_1 H_I(x)}dH_I(x)}{\mu_1 e^{\eta H_D(t_I)-\mu_1 H_I(t_I)} + \mu_1\mu_2\int_{t_I}^{t} e^{\eta H_D(x)-\mu_1 H_I(x)}dH_I(x) + \mu_2 e^{\eta H_D(t)-\mu_1 H_I(t)}}. \end{aligned}$$

$d\Lambda_{Ii}(t)$ is the hazard of subject i having incidence at time t, and $d\Lambda_{Di}(t;t_I)$ is the hazard of dying of cancer at time t, given incidence at time t_I. Note $\overline{H_*}(t)$ denotes the full path of H_* from time 0 to time t. For ease of notation, we shorten $\Theta_{Ii}\big(t;\beta,\overline{H_I}(t)\big)$ and $\Theta_{Di}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big)$ to $\Theta_{Ii}(t)$ and $\Theta_{Di}(t;t_I)$, respectively, in the rest of this paper. They can be derived through the following probabilistic argument (see Appendix C.3 for more details):

$$\mathbb{E}\{dN_{Ii}(t)|\mathcal{F}_i(t-)\} = Y_{Ii}(t)Pr\big(dN_{Ii}(t)=1|Y_{Ii}(t)=1\big) = Y_{Ii}(t)\Theta_{Ii}(t)dH_I(t),$$

and $\mathbb{E}\{dN_{Di}(t)|\mathcal{F}_i(t-)\} = Y_{Di}(t)Pr\big(dN_{Di}(t)=1|Y_{Di}(t)=1\big) = Y_{Di}(t)\Theta_{Di}(t)dH_D(t).$

4.3.2 Score Function

Define partial derivatives of Θ_{Ii} and Θ_{Di}, with respect to $\{dH_I(s)\}$, $\{dH_D(s)\}$, and $\boldsymbol{\beta}$, respectively, as

$$\begin{aligned}
&\dot{\Theta}_{Ii,dH_I(s)}\big(t;\beta,\overline{H_I}(t)\big) = \frac{\partial\Theta_{Ii}(t)}{\partial dH_I(s)} = 0,\\
&\dot{\Theta}_{Ii,dH_D(s)}\big(t;\beta,\overline{H_I}(t)\big) = 0,\\
&\dot{\Theta}_{Ii,\beta}\big(t;\beta,\overline{H_I}(t)\big) = \frac{\partial\Theta_{Ii}(t)}{\partial\beta},\\
&\dot{\Theta}_{Di,dH_I(s)}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big) = \frac{\partial\Theta_{Di}(t)}{\partial dH_I(s)},\\
&\dot{\Theta}_{Di,dH_D(s)}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big) = \frac{\partial\Theta_{Di}(t)}{\partial dH_D(s)},\\
&\dot{\Theta}_{Di,\beta}\big(t;t_I,\beta,\overline{H_I}(t),\overline{H_D}(t)\big) = \frac{\partial\Theta_{Di}(t)}{\partial\beta}.
\end{aligned}$$

As in Hu and Tsodikov (2013), for a functional $J(f)$, $f = f(x)$, the functional derivative in the above equations is defined as

$$\frac{\partial J(f)}{\partial df(s)} = \frac{\partial J(f+\epsilon g)}{\partial\epsilon}\bigg|_{\epsilon=0,g=\mathbb{1}(x>s)}.$$

We can rewrite the log-likelihood of subject i in counting process form:

$$\begin{aligned}
l_i(\beta,H_I,H_D) = \int_0^\tau \Big\{ &\Big[\log\Theta_{Ii}(t) + \log dH_I(t)\Big]dN_{Ii}(t) - Y_{Ii}(t)\Theta_{Ii}(t)dH_I(t)\\
&+\Big[\log\Theta_{Di}(t) + \log dH_D(t)\Big]dN_{Di}(t) - Y_{Di}(t)\Theta_{Di}(t)dH_D(t)\Big\}.
\end{aligned} \tag{4.11}$$

Here, τ is the duration of the study. Using l_{Ii} and l_{Di} to denote the quantities in each line of equation (4.11), and $l_I = \sum_{i=1}^n l_{Ii}$, $l_D = \sum_{i=1}^n l_{Di}$, the full log-likelihood can

be written as

$$l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} (l_{Ii} + l_{Di}) = l_I + l_D.$$

It is easy to see that l_I corresponds to the likelihood contributions from incidence, and l_D corresponds to the contributions from the subsequent time segments between incidence and death.

Applying the functional derivative to the full log-likelihood, with respect to the infinite-dimensional parameters $\{dH_I(s)\}$ and $\{dH_D(s)\}$, we can obtain the score equations for $\{dH_I(s)\}$ and $\{dH_D(s)\}$ as

$$U_{dH_I(s)} = \sum_{i=1}^{n} \left\{ \frac{dM_{Ii}(s)}{dH_I(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_I(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \tag{4.12}$$

$$U_{dH_D(s)} = \sum_{i=1}^{n} \left\{ \frac{dM_{Di}(s)}{dH_D(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_D(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \tag{4.13}$$

which are both martingales under the true model.

Taking derivative of the log-likelihood, with respect to the regression parameter β, we can have the score function for β as

$$U_\beta = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t) + \frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \tag{4.14}$$

which is also a martingale under the true model.

Set equations (4.12), (4.13), and (4.14) to zero, and solve them, can give the NPMLE $\hat{\Omega} = (\hat{\beta}, \{d\hat{H}_I\}, \{d\hat{H}_D\})$.

4.3.3 Nonparametric Maximum Likelihood Estimator

The NPMLE can be obtained using

$$(\hat{\beta}, \{d\hat{H}_I\}, \{d\hat{H}_D\}) = \underset{\beta, \{dH_I\}, \{dH_D\}}{\operatorname{argmax}} \; l(\beta, \{dH_I\}, \{dH_D\}).$$

However, it is unpleasant to directly maximize, since the nonparametric parts $\{dH_I\}$ and $\{dH_D\}$ are of infinite-dimension. Instead, we apply the profile likelihood to estimate β, $\{dH_I\}$ and $\{dH_D\}$ jointly. We first obtain the estimators of $\{dH_I\}$ and $\{dH_D\}$ with fixed β. Then replacing $\{dH_I, dH_D\}$ in the observed log-likelihood $l(\beta, \{dH_I\}, \{dH_D\})$ with $\{d\hat{H}_I(\beta), d\hat{H}_D(\beta)\}$, we have the profile log-likelihood $l_{pr} = l\big(\beta, \{d\hat{H}_I(\beta)\}, \{d\hat{H}_D(\beta)\}\big)$. Finally, the finite-dimensional parameter β is estimated by maximizing the resulting profile likelihood over β.

The key step is to obtain the estimator of $\{dH_I\}$ and $\{dH_D\}$, given β. In our model, since type of incidence is always unobserved, and in view of the simplicity of complete-data problem, we use EM algorithm to estimate the baseline hazards. For the implementation of EM, E-step is to impute the missing data, and M-step is to maximize the complete-data likelihood with imputed data plugged in. Derivation of EM algorithm for our model is shown in Appendix C.5. It gives us the score functions for dH_I and dH_D in $(k+1)th$ iteration as

$$U_{dH_I^{(k+1)}}(s) = \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - \Psi_I(s) + \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right]\theta_I(s) = 0, \tag{4.15}$$

$$U_{dH_D^{(k+1)}}(s) = \frac{dN_D(s)}{dH_D^{(k+1)}(s)} - \Psi_D(s) = 0, \tag{4.16}$$

where

$$\begin{aligned}
\Psi_I(s) &= Y_I(s)(\mu_1+\mu_2) \\
&\quad + \Delta_I Y_D(s)[1-Y_I(s)]\mu_1\mu_2 \frac{\mu_1 V(s^-, X_D) - W(s) + (1-\Delta_D)W(X_D)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}, \\
\theta_I(s) &= (1-\Delta_I)[1-Y_D(s)]\mu_1 e^{\mu_1 H_I(X_I) - \mu_1 H_I(s)} \\
&\quad + \Delta_I\mu_1\mu_2 \frac{Y_D(s)[1-Y_I(s)]W(s) + [1-Y_D(s)](1-\Delta_D)e^{\eta H_D(X_D) - \mu_1 H_I(s)}}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}, \\
\Psi_D(s) &= \Delta_I Y_D(s)[1-Y_I(s)]\eta\mu_1 \frac{W(X_I) + \mu_2 V(X_I, s^-)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}.
\end{aligned}$$

Here,

$$V(a,b) = \int_a^b e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x),$$

$$W(s) = e^{\eta H_D(s) - \mu_1 H_I(s)}.$$

Solving equations (4.15) and (4.16), we can have Breslow-type estimators

$$dH_I^{(k+1)}(s) = \frac{\sum_i dN_{Ii}(s) + \Big[\sum_i \theta_{Ii}^{(k)}(s)\Big] dH_I^{(k)}(s)}{\sum_i [\Psi_{Ii}^{(k)}(s) + \theta_{Ii}^{(k)}(s)]}, \tag{4.17}$$

$$dH_D^{(k+1)}(s) = \frac{\sum_i dN_{Di}(s)}{\sum_i \Psi_{Di}^{(k)}(s)}. \tag{4.18}$$

Equations (4.17) and (4.18) solve for $dH_I(s)$ and $dH_D(s)$ iteratively, $k = 0, 1, 2, ...$, until convergence, i.e., $dH_I^{(k+1)} \to dH_I^{(k)}$, $dH_D^{(k+1)} \to dH_D^{(k)}$. Note, at convergence, the last term in equation (4.15) disappears, and the estimating equation is the same as that obtained from observed data; equation (4.16) also takes the same form as the score function for $dH_D(s)$ from observed data. Estimators at convergence are consistent (Tsodikov (2003)).

The estimation procedure is described as follows:

Start with $\beta^{(0)} = 0$, $j = 0$.

1. Maximize the likelihood over $H_I(\beta)$ and $H_D(\beta)$, respectively, given $\beta = \beta^{(j)}$:

   (a) Set $k = 0$. Initialize $\widehat{dH}_I^{(0)}(s)$ and $\widehat{dH}_D^{(0)}(s)$ such that all jumps in the baseline hazards have equal size, respectively.

   (b) With β fixed, calculate $d\hat{H}_I^{(k+1)}(s)$ and $d\hat{H}_D^{(k+1)}(s)$ using equations (4.17) and (4.18).

(c) Repeat step (b) to update $d\hat{H}_I^{(k+1)}(s)$ and $d\hat{H}_D^{(k+1)}(s)$, until convergence

$$\left\|d\hat{H}_I^{(k+1)}(s) - d\hat{H}_I^{(k)}(s)\right\|_2 < \epsilon \quad \text{and} \quad \left\|d\hat{H}_D^{(k+1)}(s) - d\hat{H}_D^{(k)}(s)\right\|_2 < \epsilon.$$

2. Maximize the profile log-likelihood $l_{pr}(\beta) = l(\beta, \{d\hat{H}_I(\beta)\}, \{d\hat{H}_D(\beta)\})$ over β:

   (a) Calculate profile log-likelihood $l(\beta, \{d\hat{H}_I(\beta)\}, \{d\hat{H}_D(\beta)\})$ using equation (4.10).

   (b) Find $\beta^{(j+1)}$ by maximizing $l_{pr}(\beta)$ over β: $\beta^{(j+1)} = \text{argmax}_\beta\, l_{pr}(\beta)$, using conventional optimization method, e.g. Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS).

Iteratively apply steps 1-2 to estimate β, until convergence of $l_{pr}(\beta)$

$$l_{pr}(\beta^{(j+1)}) - l_{pr}(\beta^{(j)}) < \xi.$$

Note the convergence tolerance for the inner loop (EM algorithm used to estimate baseline hazards, given β) should be stricter than that for the outer loop, e.g. $\epsilon = 10^{-6}$, $\xi = 10^{-5}$.

4.4 Asymptotic Properties

We apply the empirical process (Kosorok, 2008; Van Der Vaart and Wellner, 2000) and the theory of martingale structure in counting process to build the asymptotic properties, adapted from previous work (Zeng and Lin, 2007, 2010; Chen, 2009, 2010; Hu and Tsodikov, 2014; Rice and Tsodikov, 2017).

Assuming regularity conditions hold, in the following, Theorem IV.1 and Theorem IV.2 state the consistency and weak convergence results of the NPMLE $\hat{\Omega} = (\hat{\beta}, \{d\hat{H}_I\}, \{d\hat{H}_D\})$, while Theorem IV.3 justifies the use of negative Hessian matrix

from profile log-likelihood in variance estimation. Regularity conditions and proofs are provided in Appendix C.6.

Under regularity conditions,

Theorem IV.1. *With probability 1:* $\hat{\beta}$ *converges to* β^0*;* $\hat{H}(t) = (\hat{H}_I(t), \hat{H}_D(t))$ *converges to* $H^0(t) = (H_I^0(t), H_D^0(t))$ *uniformly over the interval* $[0, \tau]$*, respectively. Here,* β^0*,* $H_I^0(t)$ *and* $H_D^0(t)$ *are the true values of* $\hat{\beta}$*,* $\hat{H}_I(t)$ *and* $\hat{H}_D(t)$*.*

Theorem IV.2. $n^{1/2}\big\{\hat{\beta}-\beta^0, \hat{H}(t)-H^0(t)\big\}$ *converges weakly to a zero-mean Gaussian process. In addition, consider a linear functional of* $\hat{\Omega}$*,*

$$n^{1/2}\bigg\{a^T(\hat{\beta}-\beta^0) + \int_0^\tau \Big[b_1(t)d\big(\hat{H}_I(t) - H_I^0(t)\big) + b_2(t)d\big(\hat{H}_D(t) - H_D^0(t)\big)\Big]\bigg\},$$

where a *is a real vector,* $b_1(t)$ *and* $b_2(t)$ *are functions with bounded total variation in* $[0, \tau]$*, evaluated at the observed incidence and death times, respectively. Let* $\Phi^T = (a^T, \{b_1(\cdot)\}^T, \{b_2(\cdot)\}^T)$*. The asymptotic variance-covariance function of the linear functional above can be consistently estimated by* $\Phi^T(\mathcal{I}_n)^{-1}\Phi$*, where* $\mathcal{I}_n$ *is the observed information matrix for* Ω*, that is,* $\mathcal{I}_n = -\dfrac{\partial^2 l_n}{\partial\Omega\partial\Omega^T}\Big|_{\Omega=\hat{\Omega}}$*, where* $l_n = n^{-1}\sum_{i=1}^n l_i$*,* l_i *is subject* i*'s observed log-likelihood defined as equation (4.11).*

Theorem IV.3. *The inverse of the negative Hessian matrix of the profile log-likelihood with respect to* β *is a consistent estimator of the limiting variance-covariance matrix of* $\hat{\beta}$*. That is,*

$$\bigg(-\frac{\partial^2 l_{pr,n}}{\partial\beta\partial\beta^T}\bigg|_{\beta=\hat{\beta}}\bigg)^{-1} \xrightarrow{p} Var\big[\sqrt{n}(\hat{\beta}-\beta^0)\big],$$

where $l_{pr,n} = n^{-1}\sum_{i=1}^n l_{pr,i}(\beta) = n^{-1}\sum_{i=1}^n l_i\big(\beta, \{d\hat{H}_I(\beta)\}, \{d\hat{H}_D(\beta)\}\big)$*.*

4.5 Simulation Studies

This section performs Monte Carlo simulations to assess the proposed methodology. The simulation settings were as follows. The baseline cumulative hazard for incidence was $H_I(t) = 0.5t^2$, and the baseline cumulative hazard for death was $H_D(t) = t$. We considered two covariates Tx and Z_1, where Tx$\sim$*Bernoulli*(0.5), and $Z_1\sim Normal(0, 1)$. β_3 evaluated the effect of Tx on the proportion of causal incidence via $\frac{\mu_1}{\mu} = e^{-\beta_3 Tx}$. The intensity of the incidence event was $\mu = e^{\beta_1 z_1+\beta_2 Tx}$, while the intensity of the terminal event was $\eta = e^{\beta_4 z_1+\beta_5 Tx}$. The true parameters were $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (-0.5, 1, 0.8, -1, 0.5)$. Censoring was simulated from the exponential distribution *Exp*(0.3), yielding 20% intermediate and 40% terminal events censored.

4.5.1 Finite-sample Properties of Parameter Estimates

We conducted simulations to study the finite-sample properties of the parameter estimates obtained. Samples of size 500 and 1000 were examined, each with 1000 replicates. Standard errors were obtained from the numerically evaluated Hessian matrix at the solution.

The simulation results are summarized in Table 4.1. The proposed estimators perform well with diminishing bias as sample size increases. The asymptotic standard errors are close to the empirical standard deviations, validating the performance of the variance estimators. The 95% coverage probabilities for all the estimators approach the 95% nominal level for both sample sizes. Note that the estimators of β_3 and β_5 slightly underperform in terms of bias, variance estimation and 95% coverage probability. This is as expected, due to the missingness of the incidence type.

Table 4.1: Simulation results using the proposed mechanistic joint model.

N	β	Truth	Bias	ASE	ESD	95% CP
500	β_1	-0.5	0.002	0.071	0.072	0.954
	β_2	1	0.003	0.077	0.076	0.963
	β_3	0.8	-0.153	0.204	0.204	0.912
	β_4	-1	0.001	0.099	0.097	0.949
	β_5	0.5	-0.091	0.141	0.146	0.877
1000	β_1	-0.5	0.000	0.048	0.050	0.964
	β_2	1	0.007	0.053	0.054	0.957
	β_3	0.8	-0.092	0.131	0.139	0.935
	β_4	-1	0.003	0.069	0.069	0.947
	β_5	0.5	-0.056	0.101	0.106	0.910

ASE: average of estimated standard errors
ESD: empirical standard deviation based on Monte Carlo estimates
95% CP: 95% coverage probability

4.5.2 Hypothesis Testing

Given the proposed estimators, we constructed the likelihood ratio test (LRT) to test the covariate effect on the terminal event. The simulation set-up was the same as before. In our scenario, the null hypothesis of no benefit of treatment on the marginal terminal event was:

$$H_0 : \ \beta_2 = \beta_3, \ \beta_5 = 0.$$

Sample size was chosen to be 1000. The simulation procedure was as follows:

1. Generate a data set under H_0.
2. Fit the models with and without the restriction, respectively, and obtain the two log-likelihoods.
3. Calculate the LRT statistic, and decide if H_0 is rejected.
4. Repeat steps 1-3 2000 times, and calculate the empirical p-value.

Based on the simulations, the 0.05 significance level is well maintained (0.051).

4.6 Application to Prostate Cancer Screening Trial Data

4.6.1 Data

The proposed method was applied to the motivating setting of the prostate cancer screening trial, testing the screening effect on cancer mortality. The data come from the PLCO trial, with patients entering the trial aged 55-74 years old. The control arm in the PLCO trial was contaminated (Vickers, 2017), as about 50% of PLCO control patients had PSA testing before enrollment, and of the remainder, close to 90% had PSA measured during the trial (Shoag et al., 2016). Thus, we needed to introduce a set of uncontaminated control data from external data to assess the screening effect. Comparing SEER with PLCO data (Pinsky et al., 2012), there was no unambiguous evidence showing a healthy volunteer effect. Therefore, a simulated subset of SEER data, with diagnosis between 1980 and 1987 before the use of PSA as a screening tool, was created to act as uncontaminated "perfect" controls.

In the combined data set, 76,674 subjects are from PLCO trial (38,335 subjects in the screening arm, and 38,339 subjects in the control arm), and 38,335 subjects are from the SEER control arm. For patients from the PLCO trial, 4418 (11.52%) in the screening arm were diagnosed with prostate cancer, and 145 (0.38%) died of it; while 4036 (10.53%) in the control arm were diagnosed and 142 (0.37%) died of it. For patients in the simulated SEER control arm, 2726 (7.11%) were diagnosed with prostate cancer, and 606 (1.58%) died of it. Maximum follow-up time was 13 years. From the data, we have information about time-to-diagnosis and time-to-death. In addition to arm (screening/control) and trial (PLCO/SEER), we are also interested in studying the age effect, since age is an important risk factor for prostate cancer. To make patients from different trials comparable, the simulated SEER data have the same age distribution as the PLCO data.

4.6.2 Results

In our proposed method, covariates entered the model via

$$\mu_1 = e^{\beta_1 Age + (\beta_2 - \beta_4) Arm + (\beta_3 - \beta_5) Trial}$$

$$\mu_2 = (1 - e^{-\beta_4 Arm + \beta_5 Trial}) e^{\beta_1 Age + \beta_2 Arm + \beta_3 Trial}$$

$$\eta = e^{\beta_6 Age + \beta_7 Arm + \beta_8 Trial}$$

The goodness of fit for the Cox regression model for marginal incidence was checked. The proportional hazards assumptions for the covariates age, arm and trial are supported, with all the three single goodness of fit p-values <0.0001. Table 4.2 shows the regression coefficient estimates for the proposed joint model. Based on the analysis of cancer diagnosis incidence, older patients have greater risk of cancer detection (β_1=0.22, HR=1.69, p-value<0.0001), and screening quickens the time-to-diagnosis (PLCO screening vs. PLCO control: β_2=0.10, HR=1.11, p-value<0.0001; PLCO control vs. SEER control: β_3=0.64, HR=1.90, p-value<0.0001). The big difference between β_2 and β_3 may be due to the contamination of the PLCO control arm, since β_2 is the log hazard ratio of comparing "systematic versus opportunistic screening" (Vickers, 2017), while β_3 is the log hazard ratio of comparing opportunistic screening versus control.

Regarding the analysis of cancer mortality given causal incidence, the risk of dying from prostate cancer significantly increases for older patients (β_6=0.19, HR=1.21, p-value<0.0001). The negative signs of β_7 and β_8 indicate that given the same occurrence of causal incidence, patients in the screening arm have a larger gap time and better survival than those in the control arm. This can be at least partly explained by the fact that with the same time of causal incidence, the screening patients tend to have better health conditions than the control patients, thus lower risk of death, even under H_0 (no benefit of screening on mortality). Yet it is difficult to separate

and interpret the screening effect from β_7 and β_8 directly.

Table 4.2: Prostate cancer screening trial analysis results.

	Parameters	Est.	SE	p-value
	β_1	0.224	0.009	<0.0001
Cancer	β_2 (PLCO scr vs. PLCO control)	0.102	0.022	<0.0001
(Diagnosis)	β_3 (PLCO control vs. SEER control)	0.642	0.025	<0.0001
Incidence	β_4	0.008	1.561	0.996
	β_5	0.504	1.367	0.712
Cancer Death	β_6	0.188	0.036	<0.0001
Given Causal	β_7 (PLCO scr vs. PLCO control)	-0.159	1.578	0.920
Incidence	β_8 (PLCO control vs. SEER control)	-1.541	1.385	0.266

The proposed model can be adopted to test the screening benefit. With this model specification, the null hypothesis of no benefit of screening on cancer mortality is

$$H_0 : \quad \beta_2 = \beta_4,\ \beta_3 = \beta_5,\ \beta_7 = \beta_8 = 0$$

Both the full model and the reduced model under H_0 were fitted to conduct the likelihood ratio test. The test statistic is 361.2 $\gg$ Chi-squared statistic with 4 df at level 0.05 = 9.488. Thus, we can reject H_0, and conclude that screening does affect cancer mortality.

To assess model fit (Dejardin et al., 2010), Figure 4.2 and 4.3 present the survival estimates for time-to-diagnosis and time-to-death, respectively, for subjects who entered the study at age 62, stratified by arm and trial. Both of the figures have plots of marginal survival functions obtained using the proposed model, which closely match the KM estimates obtained from the observed data, indicating that our proposed model works well to describe and fit the data.

4.7 Discussion

We have presented a framework to test the screening effect on cancer mortality with a mechanistic joint model of ordered events. Within the proposed model, the

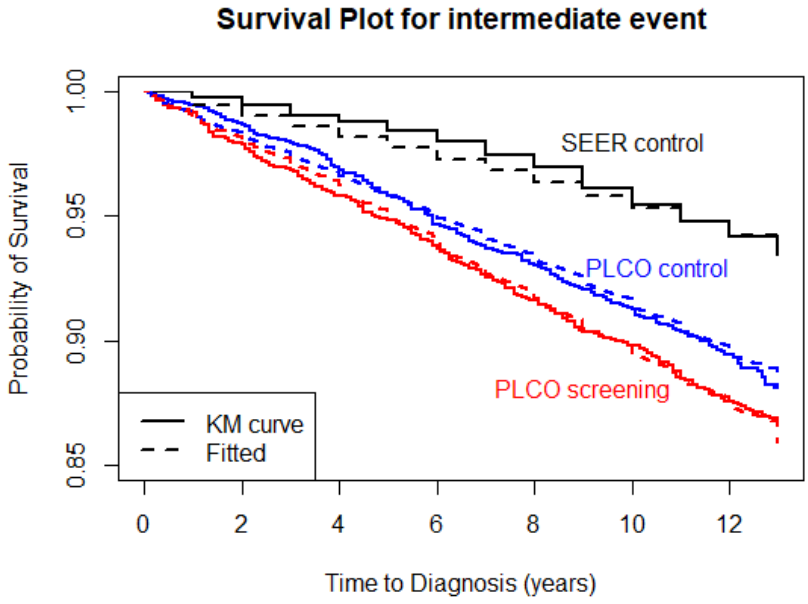

Figure 4.2: Survival estimates for time-to-diagnosis for subjects who entered the study at age 62. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

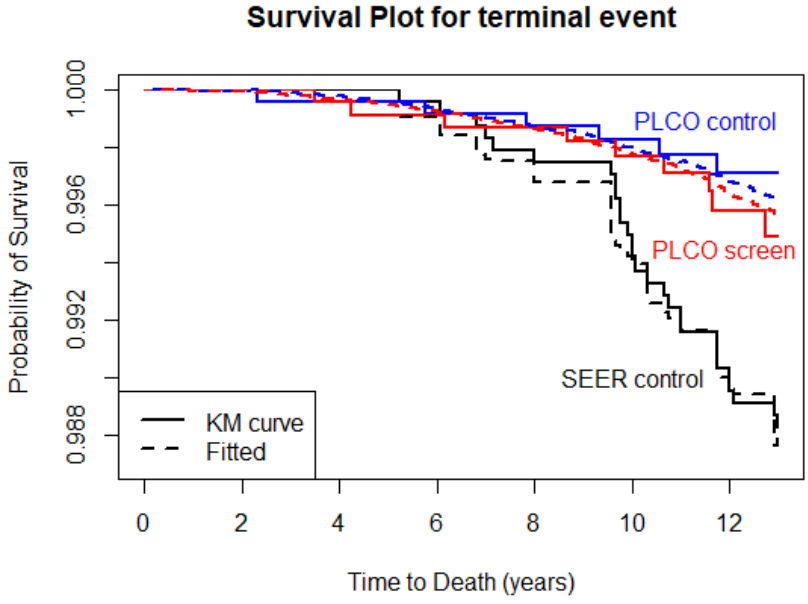

Figure 4.3: Survival estimates for time-to-death for subjects who entered the study at age 62. Proposed model (dotted lines) closely matches Kaplan-Meier (KM) estimates (solid lines).

NPMLE can be obtained by EM algorithm and profile likelihood, and its asymptotic properties established. Simulation studies indicate good finite-sample performance. Application of the proposed model to the PLCO prostate cancer data, combined with simulated SEER control data, reveals the benefit of screening on reducing cancer mortality.

The traditional joint models usually let an effect on the diagnosis propagate through the dependence structure to the terminal event, thus excluding the null hypothesis. In our formulation, we consider a partially observed disease progression process where overdiagnosis may be caused. This mechanism incorporates the null hypothesis, where screening does not affect cancer mortality, although it increases the risk of the incidence event.

Compared with the models in Chapter II and III, the proposed model can provide efficiency gains. We sampled a subset of data with 10,000 observations from the prostate cancer screening trial data, and conducted the LRT using both the model in Chapter II and the proposed mechanistic model. Both tests show significance, yet the p-value from the Chapter II model (p-value $= 1.41 \times 10^{-6}$) is twice as big as from the mechanistic model (p-value $= 7.00 \times 10^{-7}$). The mechanistic model recognizes a common cancer progression process driving incidence and mortality, thus sharing information and building a strong link between the events if the mechanism assumed is true. However, if we are uncertain about the mechanism, models in the previous chapters are more robust and may be safer choices. In other words, it is critical to balance the trade-off between robustness and efficiency gains.

CHAPTER V

Conclusion

This dissertation presents three different semiparametric joint models to test the screening effect on cancer mortality. All three models show screening benefit of prolonging cancer survival with the joint analysis of prostate cancer screening trial and population registry data. Models in Chapter II and III are under the same framework of conditional modeling approach, except the conditional incidence model is parametric in Chapter II and extended with a semiparametric form in Chapter III ; while Chapter IV proposes a mechanistic model with a latent structure. Comparing the three models, the previous two models are more robust with respect to violations of the mechanistic assumptions; the mechanistic model provides a more realistic description of a common cancer progression process driving incidence and mortality. It captures the most plausible dependence structure between the two events and endows the model parameters with a biological meaning in terms of a causal and a non-causal intermediate events. In addition, due to its latent structure, even if restricted for identifiability, the mechanistic model has more degrees of freedom, thus is more flexible. Despite the model complexity, it is more efficient when testing the null hypothesis of no screening benefit.

The current work may be extended in several directions:

- Evaluate the relative contribution of treatment advances over calendar time.

  In real data analysis, we use the PLCO trial of prostate cancer, combined with a simulated subset of SEER data with diagnosis between 1980 and 1987. In the PLCO trial, prostate cancer can be detected between 1993 and 2009, which is different from screening naive simulated SEER data before 1988. When evaluating the screening effect comparing the PLCO trial with population data, it is important to separate the effect of screening from the effect of the new treatments in later years. The question of whether new treatment contributes to better prostate cancer survival may call for additional modeling efforts. Mortality can be regressed on some surrogate of screening intensity, such as the mean lead time in the population. Model is fit to the 1993-2009 period of SEER data when new treatments are operating. Then a model prediction with zero lead time is calculated that gives the back extrapolated predicted mortality with new treatments but without screening. Comparing this prediction with observed pre-1988 mortality gives an idea of how much progress in treatment contributed to the estimated effect.

- More complex observed terminal event.

  The observed terminal event may be more complicated with several causes of death, which can potentially be related to prostate cancer. Inverse probability of censoring weighting or joint frailty modeling can be introduced to deal with this dependent censoring.

- Optimal treatment regimes.

  Consider the situation with observed endpoints of cancer recurrence and death. Initial primary treatment may affect the timing of recurrence, and also the effectiveness of recurrence treatment. Our framework in this dissertation may

be adopted to study the optimal treatment regimes in a difficult situation where a more aggressive primary treatment improves the time to recurrence, but makes the salvage treatment less effective. This problem calls for the development of a spending function characterizing the dynamic treatment regimes.

APPENDICES

APPENDIX A

Conditional Modeling of Incidence of Cancer Diagnosis on Terminal Event

A.1 Prediction of survival function for the terminal event

This section presents the derivation of conditional survival functions of terminal event given incidence information at time t^* and estimates of β and $\{dH\}$. Specifically, we are interested in $Pr(T_D > t|T_I, \Delta_1)$.

- If $t \geq t^*$,

  1. Consider a subject that has been diagnosed at t^*, i.e. $\Delta_1 = 1$

$$
\begin{aligned}
&Pr(T_D > t|T_I = t^*) = \frac{Pr(T_I = t^*, T_D > t)}{Pr(T_I = t^*)} = \frac{L_{10}(t^*, t)}{L_{10}(t^*, t^*)} \\
=&\frac{\int_t^{t_{lf}} f_{T_D}(t_D) f_{T_I|T_D}(t^*|t_D)\, dt_D + f_{T_D}(\tau) f_{T_I|T_D}(t^*|\tau)}{\int_{t^*}^{t_{lf}} f_{T_D}(t_D) f_{T_I|T_D}(t^*|t_D)\, dt_D + f_{T_D}(\tau) f_{T_I|T_D}(t^*|\tau)} \\
=&\frac{\int_t^{t_{lf}} \eta e^{-\eta H(t_D)} (\frac{t^*}{t_D})^{a-1} (1 - \frac{t^*}{t_D})^{b-1} \frac{1}{t_D} dH(t_D) + e^{-\eta H(t_{lf})} (\frac{t^*}{\tau})^{a-1} (1 - \frac{t^*}{\tau})^{b-1} \frac{1}{\tau}}{\int_{t^*}^{t_{lf}} \eta e^{-\eta H(t_D)} (\frac{t^*}{t_D})^{a-1} (1 - \frac{t^*}{t_D})^{b-1} \frac{1}{t_D} dH(t_D) + e^{-\eta H(t_{lf})} (\frac{t^*}{\tau})^{a-1} (1 - \frac{t^*}{\tau})^{b-1} \frac{1}{\tau}}
\end{aligned}
$$

2. Consider a subject that has not been diagnosed at t^*, i.e. $\Delta_1 = 0$

$$Pr(T_D > t | T_I > t^*) = \frac{Pr(T_I > t^*, T_D > t)}{Pr(T_I > t^*)} = \frac{L_{00}(t^*, t)}{L_{00}(t^*, t^*)}$$
$$= \frac{\int_t^{t_{lf}} \int_{t^*}^{t_D} f_{T_D}(t_D) f_{T_I|T_D}(t_I|t_D)\, dt_I dt_D + \int_{t^*}^{\tau} f_{T_D}(\tau) f_{T_I|T_D}(t_I|\tau)\, dt_I}{\int_{t^*}^{t_{lf}} \int_{t^*}^{t_D} f_{T_D}(t_D) f_{T_I|T_D}(t_I|t_D)\, dt_I dt_D + \int_{t^*}^{\tau} f_{T_D}(\tau) f_{T_I|T_D}(t_I|\tau)\, dt_I}$$

Here, the numerator is

$$L_{00}(t^*, t) = \int_t^{t_{lf}} \eta e^{-\eta H(t_D)} \int_{\frac{t^*}{t_D}}^{1} \left(\frac{t_I}{t_D}\right)^{a-1} \left(1 - \frac{t_I}{t_D}\right)^{b-1} d\left(\frac{t_I}{t_D}\right) dH(t_D)$$
$$+ e^{-\eta H(t_{lf})} \int_{\frac{t^*}{\tau}}^{1} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} d\left(\frac{t_I}{\tau}\right),$$

and the denominator is

$$L_{00}(t^*, t^*) = \int_{t^*}^{t_{lf}} \eta e^{-\eta H(t_D)} \int_{\frac{t^*}{t_D}}^{1} \left(\frac{t_I}{t_D}\right)^{a-1} \left(1 - \frac{t_I}{t_D}\right)^{b-1} d\left(\frac{t_I}{t_D}\right) dH(t_D)$$
$$+ e^{-\eta H(t_{lf})} \int_{\frac{t^*}{\tau}}^{1} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} d\left(\frac{t_I}{\tau}\right).$$

- If $t < t^*$,

  Since death must happen after incidence,

$$Pr(T_D > t | T_I, \Delta_1) = 1.$$

A.2 Derivation of hazard terms

For ease of notation, we suppress the subscript i for each individual subject.

1. The hazard of having incidence at time t:

$$d\Lambda_1(t) = Pr\big(dN_1(t) = 1 | Y_1(t) = 1\big) = \frac{Pr(T_I = t)}{Pr(T_I \geq t)} = \frac{L_{10}(t,t)}{L_{00}(t,t)}.$$

Here, the numerator is

$$L_{10}(t,t) = \int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} \left(\frac{t}{t_D}\right)^{a_i - 1} \left(1 - \frac{t}{t_D}\right)^{b_i - 1} \frac{1}{t_D} dH(t_D) + e^{-\eta_i H(t_{lf})} \left(\frac{t}{\tau}\right)^{a_i - 1} \left(1 - \frac{t}{\tau}\right)^{b_i - 1} \frac{1}{\tau},$$

and the denominator is

$$L_{00}(t,t) = \int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} \int_{\frac{t}{t_D}}^{1} \left(\frac{t_I}{t_D}\right)^{a_i - 1} \left(1 - \frac{t_I}{t_D}\right)^{b_i - 1} d\left(\frac{t_I}{t_D}\right) dH(t_D) + e^{-\eta_i H(t_{lf})} \int_{\frac{t}{\tau}}^{1} \left(\frac{t_I}{\tau}\right)^{a_i - 1} \left(1 - \frac{t_I}{\tau}\right)^{b_i - 1} d\left(\frac{t_I}{\tau}\right).$$

2. The hazard of dying from cancer at time t, given incidence at time t_1:

$$d\Lambda_2(t|t_1) = Pr[dN_2(t|t_1) = 1 | Y_2(t|t_1) = 1] = \frac{Pr(T_I = t_1, T_D = t)}{Pr(T_I = t_1, T_D \geq t)} = \frac{L_{11}(t_1,t)}{L_{10}(t_1,t)}$$

Here, the numerator is

$$L_{11}(t_1,t) = \eta_i e^{-\eta_i H(t)} \left(\frac{t_1}{t}\right)^{a_i - 1} \left(1 - \frac{t_1}{t}\right)^{b_i - 1} \frac{1}{t} dH(t),$$

and the denominator $L_{10}(t_1,t)$ is

$$\int_t^{t_{lf}} \eta_i e^{-\eta_i H(t_D)} (\frac{t_1}{t_D})^{a_i - 1} (1 - \frac{t_1}{t_D})^{b_i - 1} \frac{1}{t_D} dH(t_D) + e^{-\eta_i H(t_{lf})} (\frac{t_1}{\tau})^{a_i - 1} (1 - \frac{t_1}{\tau})^{b_i - 1} \frac{1}{\tau}.$$

A.3 Score functions

Based on the log-likelihood expression (2.6), we have

$$
\begin{aligned}
l_1 &= \sum_{i=1}^{n} l_{1i} = \sum_{i=1}^{n} \int_0^{\tau} dN_{1i}(t) \log d\Lambda_{1i}(t) - Y_{1i}(t) d\Lambda_{1i}(t) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \log \Theta_{1i}(t) dN_{1i}(t) - Y_{1i}(t)\Theta_{1i}(t), \\
l_2 &= \sum_{i=1}^{n} \int_0^{\tau} \int_0^{t} dN_{1i}(t_1) \Big[dN_{2i}(t) \log d\Lambda_{2i}(t;t_1) - Y_{2i}(t) d\Lambda_{2i}(t;t_1) \Big] \\
&= \sum_{i=1}^{n} \int_0^{\tau} \int_0^{t} dN_{1i}(t_1) \Big[dN_{2i}(t) \big[\log \Theta_{2i}(t;t_1) + \log dH(t) \big] - Y_{2i}(t)\Theta_{2i}(t;t_1) dH(t) \Big].
\end{aligned}
$$

Taking derivatives of l, with respect to β and $dH(s)$, respectively, we can obtain the corresponding score functions. Since $l = l_1 + l_2$, score function $U = U_1 + U_2$.

Next we derive $U_{1,(\cdot)}$. Note $(\cdot)$ can be β and $dH(s)$:

$$
\begin{aligned}
U_{1,(\cdot)} = \frac{\partial l_1}{\partial(\cdot)} &= \frac{\partial}{\partial(\cdot)} \sum_{i=1}^{n} \int_0^{\infty} \Big\{ dN_{1i}(t) \log \Theta_{1i}(t) - Y_{1i}(t)\Theta_{1i}(t) \Big\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \Big\{ \frac{\partial \log \Theta_{1i}(t)}{\partial(\cdot)} dN_{1i}(t) - Y_{1i}(t)\Theta_{1i}(t) \frac{\partial \log \Theta_{1i}(t)}{\partial(\cdot)} \Big\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \log \Theta_{1i}(t)}{\partial(\cdot)} dM_{1i}(t) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \frac{1}{\Theta_{1i}(t)} \frac{\partial \Theta_{1i}(t)}{\partial(\cdot)} dM_{1i}(t).
\end{aligned}
$$

Since $\dfrac{\partial \Theta_1(t)}{\partial dH(s)} = \mathbb{1}(t \geq s)\dfrac{\partial \Theta_1(t)}{\partial dH(s)}$, then

$$U_{1,dH(s)} = \sum_{i=1}^{n} \int_{s}^{\tau} \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t),$$
$$U_{1,\beta} = \sum_{i=1}^{n} \int_{0}^{\tau} \frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)} dM_{1i}(t).$$

Similarly, we can derive $U_{2,(\cdot)}$:

$$\begin{aligned}
U_{2,(\cdot)} = \frac{\partial l_2}{\partial(\cdot)} &= \frac{\partial}{\partial(\cdot)} \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} dN_{1i}(t_1)\Big[dN_{2i}(t) \log d\Lambda_{2i}(t;t_1) - Y_{2i}(t) d\Lambda_{2i}(t;t_1)\Big] \\
&= \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} dN_{1i}(t_1)\Big\{\frac{\partial \log d\Lambda_{2i}(t;t_1)}{\partial(\cdot)} dN_{2i}(t) - Y_{2i}(t) d\Lambda_{2i}(t;t_1) \frac{\partial \log d\Lambda_{2i}(t;t_1)}{\partial(\cdot)}\Big\} \\
&= \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} \frac{\partial \log d\Lambda_{2i}(t;t_1)}{\partial(\cdot)} dM_{2i}(t;t_1) \\
&= \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} \frac{1}{\Theta_{2i}(t;t_1) dH(t)} \frac{\partial[\Theta_{2i}(t;t_1) dH(t)]}{\partial(\cdot)} dM_{2i}(t;t_1) \\
&= \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} \frac{1}{\Theta_{2i}(t;t_1) dH(t)} \Big[dH(t) \frac{\partial \Theta_{2i}(t;t_1)}{\partial(\cdot)} + \Theta_{2i}(t;t_1) \frac{\partial dH(t)}{\partial(\cdot)}\Big] dM_{2i}(t;t_1) \\
&= \sum_{i=1}^{n} \int_{0}^{\tau} \int_{0}^{t} \Big[\frac{1}{\Theta_{2i}(t;t_1)} \frac{\partial \Theta_{2i}(t;t_1)}{\partial(\cdot)} + \frac{1}{dH(t)} \frac{\partial dH(t)}{\partial(\cdot)}\Big] dM_{2i}(t;t_1).
\end{aligned}$$

Since $\dfrac{\partial \Theta_2(t;t_1)}{\partial dH(s)} = \mathbb{1}(t \geq s)\dfrac{\partial \Theta_2(t;t_1)}{\partial dH_I(s)}$ and $\dfrac{\partial dH(t)}{\partial dH(s)} = \mathbb{1}(t = s)$, then

$$U_{2,dH(s)} = \sum_{i=1}^{n} \Big\{ \int_{0}^{s} \frac{dM_{2i}(s;t_1)}{dH(s)} + \int_{s}^{\tau} \int_{0}^{t} \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \Big\}.$$

Since $\frac{\partial dH(t)}{\partial \beta} = 0$, then

$$U_{2,\beta} = \sum_{i=1}^{n} \int_0^{\tau} \int_0^{t} \frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1).$$

Finally, since $U_{(\cdot)} = U_{1,(\cdot)} + U_{2,(\cdot)}$, score functions for $dH(s)$ and β, respectively, can be written as

$$\begin{aligned} U_{dH(s)} = \sum_{i=1}^{n} \Bigg\{ & \int_s^{\tau} \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t) \\ & + \int_s^{\tau} \int_0^{t} \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) + \int_0^{s} \frac{dM_{2i}(s;t_1)}{dH(s)} \Bigg\}, \\ U_{\beta} = \sum_{i=1}^{n} \Big\{ & \int_0^{\tau} \frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_0^{\tau} \int_0^{t} \frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \Big\}. \end{aligned}$$

A.4 EM algorithm

In our model, when the terminal event is censored, time-to-terminal T_D is missing, and EM algorithm can be adopted to estimate the baseline hazards (Tsodikov, 2003).

Conditional score function can be written as

$$U_0(s) = \frac{\partial \log L_{com}}{\partial dH(s)} = \frac{\partial \log(L_0 P_0)}{\partial dH(s)}.$$

Here, L_0 is the conditional likelihood of the incidence event, given the terminal event observed at $T_D = t_D$; and P_0 is the likelihood of the time-to-terminal T_D.

In the (k+1)th iteration,

E-step: Unconditional score function

$$U(s) = E[U_0(s)||L_0^{(k)}] = \frac{E[U_0(s)L_0^{(k)}]}{E[L_0^{(k)}]}.$$

M-step: Set $U(s) = 0$, then we can obtain $dH^{(k+1)}(s)$ to update the expression of $dH(s)$.

The rest of this section is organized as follows. First, we derive the E step for $\{dH\}$ for the three different cases respectively (both events observed, incidence observed yet death censored, and neither events observed). Then we derive the M step to maximize the likelihood with respect to $\{dH\}$, which has a closed-form expression analogous to the weighted Breslow-type estimators of Chen (2009). The following notations are introduced

$$
\begin{aligned}
U.s &= \frac{1}{B(a,b)} \eta e^{-\eta H(s)} \left(\frac{X_1}{s}\right)^{a-1} \left(1 - \frac{X_1}{s}\right)^{b-1} \frac{1}{s}, \\
U(u,v) &= \int_u^v U.s \, dH(s), \\
W.s &= \eta e^{-\eta H(s)} \int_{X_2}^{s} \frac{1}{B(a,b)} \left(\frac{t_I}{s}\right)^{a-1} \left(1 - \frac{t_I}{s}\right)^{b-1} \frac{1}{s} \, dt_I, \\
W(u,v) &= \int_u^v W.s \, dH(s), \\
V &= e^{-\eta H(t_{lf})} \frac{1}{B(a,b)} \left(\frac{X_1}{\tau}\right)^{a-1} \left(1 - \frac{X_1}{\tau}\right)^{b-1} \frac{1}{\tau}, \\
Z &= e^{-\eta H(t_{lf})} \int_{X_2}^{\tau} \frac{1}{B(a,b)} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} \frac{1}{\tau} \, dt_I.
\end{aligned}
$$

A.4.1 E step

The unconditional score functions are

$$
U_{dH}(s) = E[U_{0,dH}(s) || L_0^{(k)}] = \frac{E[U_{0,dH}(s) L_0^{(k)}]}{E[L_0^{(k)}]}
$$

In the rest of EM algorithm section, we keep iteration index for $(k+1)^{th}$ iteration when needed, and drop index for k^{th} for brevity. Note that the denominator $E[L_0^{(k)}]$

is just observed data likelihood, as in Section 2.2.3.

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$):

   Since T_D is not missing in this case,

$$U_{dH}(s) = U_{0,dH}(s) = \frac{dN_2(s)}{dH^{(k+1)}(s)} - \eta Y_2(s). \tag{A.1}$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 0$):

   The complete and observed data likelihoods differ in the two cases $X_2 \leq t_{lf}$ and $X_2 > t_{lf}$, thus we will consider them separately.

   (a) If $X_2 > t_{lf}$:

   In this case, $T_D = \tau$, the complete data likelihood can be written as

$$L_{com} = f_{T_I|T_D}(X_1|\tau) f_{T_D}(\tau) = \frac{1}{B(a,b)} \left(\frac{X_1}{\tau}\right)^{a-1} \left(1 - \frac{X_1}{\tau}\right)^{b-1} \frac{1}{\tau} e^{-\eta H(t_{lf})}.$$

   Since there is no missingness in the complete data,

$$U_{dH}(s) = U_{0,dH}(s) = -\eta Y_{lf}(s).$$

   (b) If $X_2 \leq t_{lf}$:

   First, we write out L_{com} from L_0 and P_0. Next we derive the conditional score function $U_{0,dH}(s)$ for the baseline hazard, and finally obtain the unconditional score function $U_{dH}(s)$.

   Distribution of T_D can be written as

$$P_0 = \begin{cases} \eta dH(t_D) e^{-\eta H(t_D)}, & t_D \leq t_{lf}, \\ e^{-\eta H(t_{lf})}, & t_D = \tau. \end{cases}$$

The conditional likelihood of the incidence, given $T_D = t_D$ is

$$L_0 = \begin{cases} f_{T_I|T_D}(X_1|t_D), & t_D \in (X_2, t_{lf}], \\ f_{T_I|T_D}(X_1|\tau), & t_D = \tau, \\ 0, & t_D \leq X_2. \end{cases}$$

Combining L_0 and P_0, we can obtain the complete-data likelihood as

$$L_{com} = \begin{cases} \frac{1}{B(a,b)} \left(\frac{X_1}{t_D}\right)^{a-1} \left(1 - \frac{X_1}{t_D}\right)^{b-1} \frac{1}{t_D} \eta dH(t_D) e^{-\eta H(t_D)}, & t_D \leq t_{lf}, \\ \frac{1}{B(a,b)} \left(\frac{X_1}{\tau}\right)^{a-1} \left(1 - \frac{X_1}{\tau}\right)^{b-1} \frac{1}{\tau} e^{-\eta H(t_{lf})}, & t_D = \tau. \end{cases}$$

Take derivative of L_{com} w.r.t $\{dH(s)\}$ to obtain the conditional score function as

$$U_{0,dH}(s) = \begin{cases} \frac{\mathbb{1}(T_D = s)}{dH(s)} - \eta \mathbb{1}(T_D \geq s), & T_D \in (X_2, t_{lf}] \\ -\eta \mathbb{1}(t_{lf} \geq s), & T_D = \tau. \end{cases}$$

Since $U_{0,dH}(s)$ depends on s, we consider the calculation of the unconditional score function $U_{dH}(s)$ in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf}$, and $s > t_{lf}$, respectively.

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  To calculate $U_{dH}(s)$, first we have $U_{0,dH}(s) = -\eta$.

  Since $U_{0,dH}(s)$ does not depend on the missing variable,

  $$E[U_{0,dH}(s) L_0] = U_{0,dH}(s) E[L_0] = -\eta E[L_0].$$

  Thus, $U_{dH}(s) = \dfrac{E[U_{0,dH}(s) L_0]}{E[L_0]} = -\eta.$

- $X_2 < s \leq t_{lf}$, i.e. $Y_2(s) = 0$, $Y_{lf}(s) = 1$

To calculate $U_{dH}(s)$, first

$$U_{0,dH}(s) = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH^{(k+1)}(s)} - \eta \mathbb{1}(T_D \geq s), & T_D \in (X_2, t_{lf}], \\ -\eta, & T_D = \tau. \end{cases}$$

Then the numerator

$$E[U_{0,dH}(s)L_0] = \frac{E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]}{dH^{(k+1)}(s)} - \eta E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] - \eta E[\mathbb{1}(T_D = \tau)L_0].$$

We can derive the term $E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]$ as

$$\begin{aligned} &E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] \\ =&\frac{1}{B(a,b)} \left(\frac{X_1}{s}\right)^{a-1} \left(1 - \frac{X_1}{s}\right)^{b-1} \frac{1}{s} \eta dH(s) e^{-\eta H(s)} = U.s dH(s). \end{aligned}$$

The term $E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]$ can be derived as

$$\begin{aligned} &E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] \\ =&\int_{s^-}^{t_{lf}} \frac{1}{B(a,b)} \left(\frac{X_1}{t_D}\right)^{a-1} \left(1 - \frac{X_1}{t_D}\right)^{b-1} \frac{1}{t_D} \eta e^{-\eta H(t_D)} dH(t_D) = U(s^-, t_{lf}). \end{aligned}$$

The term $E[\mathbb{1}(T_D = \tau)L_0]$ can be derived as

$$E[\mathbb{1}(T_D = \tau)L_0] = \frac{1}{B(a,b)} \left(\frac{X_1}{\tau}\right)^{a-1} \left(1 - \frac{X_1}{\tau}\right)^{b-1} \frac{1}{\tau} \eta e^{-\eta H(t_{lf})} = V.$$

Then, the numerator can be calculated as

$$E[U_{0,dH}(s)L_0] = \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right] U.s - \left[\eta U(s^-, t_{lf}) + \eta V - U.s\right].$$

Note the denominator $E[L_0^{(k)}]$ is the observed likelihood L_{10}, which can be rewritten as $U(X_2, t_{lf}) + V$. Thus, the unconditional score function for $dH(s)$ is

$$U_{dH}(s) = \frac{\left[\dfrac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right] U.s - \left[\eta U(s^-, t_{lf}) + \eta V - U.s\right]}{U(X_2, t_{lf}) + V}.$$

- $s > t_{lf}$, i.e. $Y_{lf}(s) = 0$

  Since $U_{0,dH}(s) = 0$, the unconditional score function is also zero.

Combining these results, the contribution of a subject, with intermediate event observed at X_1 yet terminal event censored at X_2, to the unconditional score functions, when $s \leq t_{lf}$, can be written as

$$U_{dH}(s) = \mathbb{1}(X_2 \leq t_{lf}) \Bigg\{ \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right] \big(1 - Y_2(s)\big) \frac{U.s}{U(X_2, t_{lf}) + V} - \left[Y_2(s)\eta + \big(1 - Y_2(s)\big) \frac{\eta U(s^-, t_{lf}) + \eta V - U.s}{U(X_2, t_{lf}) + V} \right] \Bigg\} - \mathbb{1}(X_2 > t_{lf})\eta. \quad \text{(A.2)}$$

Note, $U_{dH}(s) = 0$ when $s > t_{lf}$, indicating that there is NO jump in $\{dH(s)\}$ when $s > t_{lf}$.

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = 0$, $\Delta_2 = 0$):

   We now proceed through the same steps as for the case $\Delta_1 = 1$, $\Delta_2 = 0$.

   (a) If $X_2 > t_{lf}$:

In this case, $T_D = \tau$, the complete data likelihood is

$$\begin{aligned} L_{com} &= f_{T_D}(\tau) \int_{X_2}^{\tau} f_{T_I|T_D}(t_I|\tau) dt_I \\ &= e^{\eta H(t_{lf})} \int_{X_2}^{\tau} \frac{1}{B(a,b)} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} \frac{1}{\tau} dt_I. \end{aligned}$$

There is no missingness for the complete data, thus

$$U_{dH}(s) = U_{0,dH}(s) = -\eta Y_{lf}(s).$$

(b) If $X_2 \leq t_{lf}$:

First, we derive L_{com} from L_0 and P_0. P_0 keeps the same distribution as before, while

$$L_0 = \begin{cases} \int_{X_2}^{t_D} f_{T_I|T_D}(t_I|t_D) dt_I, & t_D \in (X_2, t_{lf}], \\ \int_{X_2}^{\tau} f_{T_I|T_D}(t_I|\tau) dt_I, & t_D = \tau, \\ 0, & t_D \leq X_2. \end{cases}$$

Combining L_0 and P_0, the complete-data likelihood can be written as

$$L_{com} = \begin{cases} \eta dH(t_D) e^{-\eta H(t_D)} \int_{X_2}^{t_D} \frac{1}{B(a,b)} (\frac{t_I}{t_D})^{a-1} (1 - \frac{t_I}{t_D})^{b-1} \frac{1}{t_D} dt_I, & t_D \leq t_{lf}, \\ e^{-\eta H(t_{lf})} \int_{X_2}^{\tau} \frac{1}{B(a,b)} (\frac{t_I}{\tau})^{a-1} (1 - \frac{t_I}{\tau})^{b-1} \frac{1}{\tau} dt_I, & t_D = \tau. \end{cases}$$

Then, take derivative of L_{com} w.r.t $\{dH(s)\}$, we can obtain the conditional score function as

$$U_{0,dH}(s) = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH(s)} - \eta \mathbb{1}(T_D \geq s), & T_D \in (X_2, t_{lf}] \\ -\eta \mathbb{1}(t_{lf} \geq s), & T_D = \tau. \end{cases}$$

Next, we calculate $U_{dH}(s)$ in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf}$, and $s > t_{lf}$, respectively.

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  To calculate $U_{dH}(s)$, first we have $U_{0,dH}(s) = -\eta$.

  $U_{0,dH}(s)$ does not depend on the missing variable, then

  $$E[U_{0,dH}(s)L_0] = U_{0,dH}(s)E[L_0] = -\eta E[L_0].$$

  Thus, $U_{dH}(s) = \dfrac{E[U_{0,dH}(s)L_0]}{E[L_0]} = -\eta$.

- $X_2 < s \leq t_{lf}$, i.e. $Y_2(s) = 0$, $Y_{lf}(s) = 1$

  To calculate $U_{dH}(s)$, first

  $$U_{0,dH}(s) = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH^{(k+1)}(s)} - \eta \mathbb{1}(T_D \geq s), & T_D \in (X_2, t_{lf}], \\ -\eta, & T_D = \tau. \end{cases}$$

  Then the numerator

  $$E[U_{0,dH}(s)L_0] = \frac{E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]}{dH^{(k+1)}(s)} - \eta E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] - \eta E[\mathbb{1}(T_D = \tau)L_0].$$

  We can derive the term $E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]$ as

  $$\begin{aligned} &E[\mathbb{1}(T_D = s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] \\ =&\eta dH(s)e^{-\eta H(s)} \int_{X_2}^{s} \frac{1}{B(a,b)} (\frac{t_I}{s})^{a-1}(1 - \frac{t_I}{s})^{b-1}\frac{1}{s}dt_I = W.sdH(s). \end{aligned}$$

The term $E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0]$ can be derived as

$$\begin{aligned}
&E[\mathbb{1}(T_D \geq s)\mathbb{1}(X_2 < T_D \leq t_{lf})L_0] \\
=&\int_{s^-}^{t_{lf}} \int_{X_2}^{t_D} \frac{\eta}{B(a,b)} \left(\frac{t_I}{t_D}\right)^{a-1} \left(1 - \frac{t_I}{t_D}\right)^{b-1} \frac{1}{t_D} e^{-\eta H(t_D)} dt_I dH(t_D) \\
=&W(s^-, t_{lf}).
\end{aligned}$$

The term $E[\mathbb{1}(T_D = \tau)L_0]$ can be derived as

$$e^{-\eta H(t_{lf})} \int_{X_2}^{\tau} \frac{1}{B(a,b)} \left(\frac{t_I}{\tau}\right)^{a-1} \left(1 - \frac{t_I}{\tau}\right)^{b-1} \frac{1}{\tau} dt_I = Z.$$

Then, the numerator can be calculated as

$$E[U_{0,dH}(s)L_0] = \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right] W.s - \left[\eta W(s^-, t_{lf}) + \eta Z - W.s\right].$$

Note the denominator $E[L_0^{(k)}]$ is the observed likelihood L_{00}, which can be rewritten as $W(X_2, t_{lf}) + Z$. Thus, the unconditional score function for $dH(s)$ is

$$U_{dH}(s) = \frac{\left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right] W.s - \left[\eta W(s^-, t_{lf}) + \eta Z - W.s\right]}{W(X_2, t_{lf}) + Z}.$$

- $s > t_{lf}$, i.e. $Y_{lf}(s) = 0$

  Since $U_{0,dH}(s) = 0$, the unconditional score function is also zero.

Combining these results, the contribution of a subject, censored at X_2 before any event is observed, to the unconditional score functions, when $s \leq t_{lf}$, can

be written as

$$U_{dH}(s) = \mathbb{1}(X_2 \le t_{lf})\Bigg\{\Big[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\Big]\big(1 - Y_2(s)\big)\frac{W.s}{W(X_2, t_{lf}) + Z} - \Big[Y_2(s)\eta + \big(1 - Y_2(s)\big)\frac{\eta W(s^-, t_{lf}) + \eta Z - U.s}{W(X_2, t_{lf}) + Z}\Big]\Bigg\} - \mathbb{1}(X_2 > t_{lf})\eta. \quad \text{(A.3)}$$

Note, $U_{dH}(s) = 0$ when $s > t_{lf}$, indicating that there is NO jump in $\{dH(s)\}$ when $s > t_{lf}$.

Combining equations (A.1), (A.2), and (A.3), we have for the unconditional score functions

$$U_{dH^{(j+1)}}(s) = \frac{dN_2(s)}{dH^{(j+1)}(s)} - \Psi^{(j)}(s) + \Big[\frac{dH^{(j)}(s)}{dH^{(j+1)}(s)} - 1\Big]\theta^{(j)}(s),$$

where

$$\Psi(s) = \mathbf{1}(X_2 \le t_{lf})\Bigg\{Y_2(s)\eta + [1 - Y_2(s)](1 - \Delta_2)\big[\Delta_1\frac{\eta U(s^-, t_{lf}) + \eta V - U.s}{U(X_2, t_{lf}) + V} + (1 - \Delta_1)\frac{\eta W(s^-, t_{lf}) + \eta Z - W.s}{W(X_2, t_{lf}) + Z}\big]\Bigg\} + \mathbf{1}(X_2 > t_{lf})\eta,$$

$$\theta(s) = \mathbf{1}(X_2 \le t_{lf})[1 - Y_2(s)](1 - \Delta_2)\Bigg\{\Delta_1\frac{U.s}{U(X_2, t_{lf}) + V} + (1 - \Delta_1)\frac{W.s}{W(X_2, t_{lf}) + Z}\Bigg\}.$$

A.4.2 M step

Suppose we have n independent subjects with observed data $(X_{1i}, \Delta_{1i}, X_{2i}, \Delta_{2i})$, $i = 1, 2, ..., n$. The estimator of $dH^{(k+1)}(s)$ can be derived by solving $\sum_{i=1}^{n} U_{i,dH}(s) = 0$. Now we want to solve

$$\sum_{i=1}^{n}\Bigg\{\frac{dN_{2i}(s)}{dH^{(k+1)}(s)} - \Psi_i(s) + \Big[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\Big]\theta_i(s)\Bigg\} = 0,$$

which gives

$$dH^{(k+1)}(s) = \frac{\sum_i dN_{2i}(s) + \Big[\sum_i \theta_i^{(k)}(s)\Big] dH^{(k)}(s)}{\sum_i [\Psi_i^{(k)}(s) + \theta_i^{(k)}(s)]}.$$

The above equation solves for $dH(s)$ iteratively until convergence, and the estimator at convergence, $\widehat{dH}(s)$, is a consistent estimator of $dH(s)$ (Tsodikov (2003)).

A.5 Asymptotic properties

This section presents the technical details of the asymptotic properties of the NPMLE $\hat{\Omega} = (\hat{\beta}, \{d\hat{H}\})$, adapted from Hu and Tsodikov (2014), Supplementary Materials C and Rice and Tsodikov (2017), Appendix F. To establish the asymptotic properties, we assume the following conditions are satisfied (Fleming and Harrington (2011), p289-p290):

1. The true values of baseline hazards H^0 are strictly increasing and differentiable. The true parameter set Ω^0 is in the interior of the compact convex set $\mathcal{H}$.

2. With probability 1, $P[Y_2(t)|z] > 0$, $P[\Delta_2 = 0, X_2 = \tau|z] > 0$. The at risk set will not shrink to zero.

3. The Hessian matrix $\mathcal{I}_n$ evaluated at the true value $\Omega^0 = (\beta^0, \{dH^0\})$ is positive definite, and converges in probability to a deterministic and invertible operator $\mathcal{I}^0$.

4. The model is identifiable such that

$$\{\Lambda\} = \{\Lambda^0\} \quad \text{uniformly over } \Omega \quad \Rightarrow \quad \Omega = \Omega^0.$$

A.5.1 Proof of Theorem II.1

To prove consistency: $|\hat{\Omega} - \Omega^0| \xrightarrow{p} 0$, based on Theorom 2.12 of Kosorok (2008), in addition to the regularity conditions provided above, another three conditions need to be verified:

(a) $l_n(\hat{\Omega}_n) = \sup_{\Omega \in \mathcal{H}} l_n(\Omega) - o_p(1)$

(b) Identifiability condition 2: For any sequence $\Omega_n \in \mathcal{H}$,

$$\liminf_{n\to\infty} l(\Omega_n) \geq l(\Omega^0) \quad \Rightarrow \quad \|\Omega_n - \Omega^0\| \xrightarrow{p} 0.$$

(c) Uniform convergence condition: $\sup_{\Omega \in \mathcal{H}} |l_n(\Omega) - l(\Omega)| \xrightarrow{p} 0$.

Note, $\hat{\Omega}_n$ here is the maximum likelihood estimator of Ω, which is $\hat{\Omega}$ in our previous notation.

We verify the three conditions as follows:

1. To verify condition (a), $\hat{\Omega}_n$ is the maximum estimator of Ω, thus condition (a) is satisfied.

2. To verify condition (b), by Lemma 14.3 of Kosorok (2008), we just need to prove that $l(\Omega)$ is upper semicontinuous with a unique maximum at Ω^0.

   The model is characterized by defining the hazard functions of $d\Lambda_1(t)$ and $d\Lambda_2(t)$, which are both functions of Ω. Let $F_1(t)$ and $F_2(t)$ be the cumulative density functions for incidence and death, subject to censoring; and let $S_1(t)$ and $S_2(t)$ be the survival functions for observed incidence and death. $F_1^0(t)$, $F_2^0(t)$, $S_1^0(t)$ and $S_2^0(t)$ denote the corresponding true functions. Note that $dF_1(t) = S_1(t)d\Lambda_1(t)$ and $dF_2(t) = S_2(t)d\Lambda_2(t)$. The true log-likelihood

can be written as

$$l(\Omega) = E \int_0^\tau \Big\{ \log d\Lambda_1(t) dF_1^0(t) - S_1^0(t) d\Lambda_1(t) + \log d\Lambda_2(t) dF_2^0(t) - S_2^0(t) d\Lambda_2(t) \Big\},$$

where the expectation is taken over covariate $\mathbf{z}$.

Now consider the negative Kullback-Leibler distance,

$$\begin{aligned} D &= l(\Omega) - l(\Omega^0) \\ &= E \int_0^\tau \Big\{ dF_1^0(t)[\log d\Lambda_1(t) - \log d\Lambda_1^0(t)] - S_1^0(t)[d\Lambda_1(t) - d\Lambda_1^0(t)] \\ &\qquad + dF_2^0(t)[\log d\Lambda_2(t) - \log d\Lambda_2^0(t)] - S_2^0(t)[d\Lambda_2(t) - d\Lambda_2^0(t)] \Big\}. \end{aligned}$$

Since

$$\begin{aligned} & dF^0(t)[\log d\Lambda(t) - \log d\Lambda^0(t)] - S^0(t)[d\Lambda(t) - d\Lambda^0(t)] \\ &= dF^0(t) \log \frac{d\Lambda(t)}{d\Lambda^0(t)} - \frac{dF^0(t)}{d\Lambda^0(t)}[d\Lambda(t) - d\Lambda^0(t)] \\ &= dF^0(t) \Big[\log \frac{d\Lambda(t)}{d\Lambda^0(t)} - \frac{d\Lambda(t)}{d\Lambda^0(t)} + 1 \Big], \end{aligned}$$

then

$$\begin{aligned} D &= E \int_0^\tau \Big\{ dF_1^0(t) \big[\log \frac{d\Lambda_1(t)}{d\Lambda_1^0(t)} - \frac{d\Lambda_1(t)}{d\Lambda_1^0(t)} + 1 \big] + dF_2^0(t) \big[\log \frac{d\Lambda_2(t)}{d\Lambda_2^0(t)} - \frac{d\Lambda_2(t)}{d\Lambda_2^0(t)} + 1 \big] \Big\} \\ &= E \int_0^\tau \left\{ \rho\left(\frac{d\Lambda_1(t)}{d\Lambda_1^0(t)} \right) dF_1^0(t) + \rho\left(\frac{d\Lambda_2(t)}{d\Lambda_2^0(t)} \right) dF_2^0(t) \right\}, \end{aligned}$$

where $\rho(x) = \log x - x + 1$ is a non-positive convex function, with a unique maximizer at $x = 1$. Therefore, D has a unique maximum at $\dfrac{d\Lambda_1(t)}{d\Lambda_1^0(t)} = 1$ and $\dfrac{d\Lambda_2(t)}{d\Lambda_2^0(t)} = 1$, i.e., $d\Lambda_1(t) = d\Lambda_1^0(t)$ and $d\Lambda_2(t) = d\Lambda_2^0(t)$.

Given regularity condition 4, under the identifiable model, D has a unique maximum at Ω^0. Since maximizing D is equivalent to maximizing $l(\Omega)$, we can conclude that $l(\Omega)$ has a unique maximum at Ω^0. We also know $l(\Omega)$ is upper semicontinuous, thus condition (b) holds.

3. To verify condition (c), given regularity condition 1, Ω is in the class of functions of bounded variation with integrable envelope, so $H(t)$ is bounded. Therefore, $\mathcal{H}$ is a Glivenko-Cantelli class. Then, since the functionals Λ and $l(\Omega)$ are continuous, and the envelope function is integrable, then by the preservation of Glivenko-Cantelli theorem (Van Der Vaart and Wellner (2000)), the integrand in $l(\Omega)$ is also Glivenko-Cantelli. Therefore, we apply the uniform law of large numbers for the empirical process, such that

$$\sup_{\Omega \in \mathcal{H}} |l_n(\Omega) - l(\Omega)| \xrightarrow{p} 0.$$

A.5.2 Martingale Representation of Score Functions

In this part, we justify that the score functions for $H(t)$ and β are all martingales under the true model.

We first consider the score function $U_{H(x)}$. Based on score function (2.7), we integrate the expression over $dH(s)$, normalize it by $1/n$, and obtain the normalized score function for $H(x)$ as

$$U_{H(x)} = \frac{1}{n}\sum_{i=1}^{n} \int_0^x \Big\{ \int_s^\tau \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_0^s \frac{dM_{2i}(s;t_1)}{dH(s)} + \int_s^\tau \int_0^t \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1) \Big\}.$$

Exchange the integrals, then we have

$$U_{H(x)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^\tau \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)}H(x\wedge t)dM_{1i}(t) + \int_0^\tau\int_0^t (t\le x)dM_{2i}(t;t_1) + \int_0^\tau\int_0^t \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)}H(x\wedge t)dM_{2i}(t;t_1)\Big\}.$$

Similar as the proof in (Hu and Tsodikov (2014), Supplementary Materials B), Let $\epsilon_1(t,x) = \frac{\dot{\Theta}_{1i,dH(s)}(t)}{\Theta_{1i}(t)}H(x\wedge t)$, $\epsilon_2(t,x) = \frac{\dot{\Theta}_{2i,dH(s)}(t;t_1)}{\Theta_{2i}(t;t_1)}H(x\wedge t)$, then

$$U_{H(x)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^\tau\int_0^t (t\le x)dM_{2i}(t;t_1) + \int_0^\tau \epsilon_{1i}(t,x)dM_{1i}(t) + \int_0^\tau\int_0^t \epsilon_{2i}(t,x)dM_{2i}(t;t_1)\Big\}.$$

Consider the increment of $U_{H(x)}$ over x,

$$\begin{aligned} dU_H(x) &= U_{H(x+dx)} - U_{H(x)} \\ &= \int_0^x dM_2(x;t_1) + \int_0^\tau \frac{\partial\epsilon_1(t,x)}{\partial x}dxdM_1(t) + \int_0^\tau\int_0^t \frac{\partial\epsilon_2(t,x)}{\partial x}dxdM_2(t;t_1). \end{aligned}$$

Take the expectation conditional on filtration $\mathcal{F}(x^-)$,

$$\begin{aligned} E[dU_H(x)|\mathcal{F}(x^-)] = E[\int_0^x dM_2(x;t_1)|\mathcal{F}(x^-)] &+ E\Big[\int_0^\tau \frac{\partial\epsilon_1(t,x)}{\partial x}dxdM_1(t)|\mathcal{F}(x^-)\Big] \\ &+ E\Big[\int_0^\tau\int_0^t \frac{\partial\epsilon_2(t,x)}{\partial x}dxdM_2(t|t_1)|\mathcal{F}(x^-)\Big]. \end{aligned}$$

The first term is 0, based on the martingale property of $M_2(x)$. $\epsilon_1(t,x)$ depends on t when $t < x$, so $\frac{\partial\epsilon_1(t,x)}{\partial x} = 0$ when $t < x$. $E[dM_1(t)|\mathcal{F}(x^-)] = 0$, if $t \ge x^-$, based on

the martingale property of M_1. So

$$
\begin{aligned}
&E\Big[\int_0^{\tau} \frac{\partial \epsilon_1(t,x)}{\partial x} dx dM_1(t) | \mathcal{F}(x^-)\Big] \\
=&\int_0^{x^-} E\Big[\frac{\partial \epsilon_1(t,x)}{\partial x} dx dM_1(t) | \mathcal{F}(x^-)\Big] + \int_{x^-}^{\tau} E\Big[\frac{\partial \epsilon_1(t,x)}{\partial x} dx dM_1(t) | \mathcal{F}(x^-)\Big] \\
=&0 + \int_{x^-}^{\tau} \frac{\partial \epsilon_1(t,x)}{\partial x} dx E[dM_1(t) | \mathcal{F}(x^-)] = 0.
\end{aligned}
$$

Similarly, $E\Big[\int_0^{\tau}\int_0^{t} \frac{\partial \epsilon_2(t,x)}{\partial x} dx dM_2(t|t_1) | \mathcal{F}(x^-)\Big] = 0$. Thus, $E[dU_H(x) | \mathcal{F}(x^-)] = 0$, and $U_H(x)$ is a martingale under the true model.

In terms of the score function for β, as expressed in equation (2.8), and after normalization by $\frac{1}{n}$, U_β becomes

$$
U_\beta = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\int_0^{\tau} \frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)} dM_{1i}(t) + \int_0^{\tau}\int_0^{t} \frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)} dM_{2i}(t;t_1)\Bigg\}.
$$

Since $\frac{\dot{\Theta}_{1i,\beta}(t)}{\Theta_{1i}(t)}$ and $\frac{\dot{\Theta}_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)}$ are both predictable, so U_β, which is the linear transformation of these two terms, is also a martingale under the true model.

A.5.3 Proof of Theorem II.2

We prove Theorem II.2 in two steps. Suppose $U(\Omega) = (U_\beta, U_{H(t)})^T$ is the set of score functions for parameter set Ω. First, we prove the weak convergence of the score functions at the true parameter $n^{1/2}U(\Omega^0)$ by Martingale Central Limit Theorem (MCLT). Then, we seek for the relationship between $n^{1/2}(\hat{\Omega} - \Omega^0)$ and $n^{1/2}U(\Omega^0)$, to obtain the weak convergence of the NPMLE $\hat{\Omega}$.

Based on the martingale representation of $U(\Omega^0)$ in Appendix A.5.2, and the fact that N_{1i} and N_{2i}, $i = 1, 2, ..., n$, are orthogonal, it follows that $n^{1/2}U(\Omega^0)$ converges

weakly to a zero-mean Gaussian process with its variance-covariance function characterized by $\sigma^2_\beta(\Omega^0)$, $\sigma^2_H(x,y;\Omega^0)$ and $\sigma^2_{\beta,H(x)}(x;\Omega^0)$ as derived below.

The predictable variation process for score function $n^{1/2}U_\beta$ is

$$\begin{aligned}
&< n^{1/2}U_\beta > \\
&= n\frac{1}{n^2}\sum_{i=1}^{n}\Big\{\int_0^\tau \frac{\dot\Theta^2_{1i,\beta}(x)}{\Theta^2_{1i}(x)}Y_{1i}(t)\Theta_{1i}(x) + \int_0^\tau\int_0^t \frac{\dot\Theta^2_{2i,\beta}(t;t_1)}{\Theta^2_{2i}(t;t_1)}\Theta_{2i}(t;t_1)Y_{2i}(t)dH(t)\Big\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^\tau Y_{1i}(t)\frac{\dot\Theta^2_{1i,\beta}(t)}{\Theta_{1i}(t)} + \int_0^\tau\int_0^t \frac{\dot\Theta^2_{2i,\beta}(t;t_1)}{\Theta_{2i}(t;t_1)}Y_{2i}(t)dH(t)\Big\} \\
&\xrightarrow{p} \int_0^\tau P(T_1\ge t)\frac{\dot\Theta^2_{1,\beta}(t)}{\Theta_1(t)} + \int_0^\tau\int_0^t \frac{\dot\Theta^2_{2,\beta}(t)}{\Theta_2(t)}P(T_2\ge t)dH(t).
\end{aligned}$$

Thus, as $n\to\infty$, $n^{1/2}U_\beta$ converges weakly to a zero-mean Gaussian process with covariance function

$$\sigma^2_\beta(\Omega^0) = \int_0^\tau P(T_1\ge t)\frac{\dot\Theta^2_{1,\beta}(t)}{\Theta_1(t)} + \int_0^\tau\int_0^t \frac{\dot\Theta^2_{2,\beta}(t)}{\Theta_2(t)}P(T_2\ge t)dH(t).$$

Similarly, as $n\to\infty$,

$n^{1/2}U_H$ converges weakly to a zero-mean Gaussian process with covariance function

$$\begin{aligned}
\sigma^2_H(x,y;\Omega^0) = &\int_0^\tau \epsilon_1(t,x)\epsilon_1(t,y)P(T_1\ge t)\Theta_1(t) \\
&+ \int_0^\tau\int_0^t \Big[\mathbb{1}(t\le x)\mathbb{1}(t\le y) + \epsilon_2(t,x)\epsilon_2(t,y)\Big]P(T_2\ge t)\Theta_2(t;t_1)dH(t),
\end{aligned}$$

for $x,y\in[0,\tau]$;

$n^{1/2}U\big(\beta,H(x)\big)$ converges weakly to a zero-mean Gaussian process with covariance

function

$$\sigma^2_{\beta,H(x)}(x;\Omega^0) = \int_0^\tau P(T_1 \geq t)\epsilon_1(t,x)\dot{\Theta}_{1,\beta}(t) + \int_0^\tau \int_0^t \Big[\mathbb{1}(t \leq x) + \epsilon_2(t,x)\Big]\dot{\Theta}_{2,\beta}(t;t_1)P(T_2 \geq t)dH(t).$$

Let the normalized log-likelihood $\frac{1}{n}\sum_{i=1}^{n} l_i$ converges in probability to l_∞, and $U_\infty = \Big(\frac{\partial l_\infty}{\partial \beta}, \frac{\partial l_\infty}{\partial dH(t)}\Big)^T$. Define a linear information operator $\mathcal{I}_\infty$ as

$$\mathcal{I}_\infty(t,s) = -\frac{\partial U_\infty}{\partial \Omega}\Big|_{\Omega=\Omega^0} = -\begin{bmatrix} \frac{\partial^2 l_\infty}{\partial\beta\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)} \\ \frac{\partial^2 l_\infty}{\partial dH(t)\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)} \end{bmatrix}_{\Omega=\Omega^0},$$

and the operator $\mathcal{I}_\infty$ acts on an arbitrary vector-function element $\Omega_s = \big(\beta, dH(s)\big)^T$ as follows

$$\mathcal{I}_\infty(t,s)\Omega_s = -\begin{bmatrix} \frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)}\big|_{\Omega=\Omega^0}\beta + \int_0^\tau \frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)}\big|_{\Omega=\Omega^0} dH(s) \\ \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)}\big|_{\Omega=\Omega^0}\beta + \int_0^\tau \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)}\big|_{\Omega=\Omega^0}(s). \end{bmatrix}$$

With Taylor expansion, expand $U(\hat{\Omega})$ at the true parameter Ω^0, we have

$$\begin{aligned} & U(\hat{\Omega}) = 0 = U(\Omega^0) - \mathcal{I}_\infty(t,s)(\hat{\Omega} - \Omega^0) + o_p(1) \\ \Rightarrow \quad & \mathcal{I}_\infty(t,s)n^{1/2}(\hat{\Omega} - \Omega^0) = n^{1/2}U(\Omega^0) + o_p(1). \end{aligned} \tag{A.4}$$

Assume that the Fredholm operator expressed by the kernel $\mathcal{I}_\infty$ of the Fredholm integral equations (A.4) of the first kind is square integrable, and the equation $\mathcal{I}_\infty\Omega = 0$ has only the trivial solution $\Omega = 0$. Then based on Theorem 3.3.1 of Vaart and Wellner (1996), equations (A.4) has the unique solution, and there exists the inverse

information operator $\mathcal{I}_\infty^{-1}(t,s)$ such that

$$n^{1/2}(\hat{\Omega}-\Omega^0)=\mathcal{I}_\infty^{-1}(t,s)n^{1/2}U(\Omega^0)+o_p(1).$$

Take differentiation of the equation $E[U(\Omega^0)]=0$ with respect to Ω at the true parameter Ω^0, we have the equivalence between $\mathcal{I}_\infty$ represented by the log-likelihood second derivative, and

$$\mathcal{I}_\infty(t,s)=\begin{bmatrix} \frac{\partial l_\infty}{\partial\beta}\frac{\partial l_\infty}{\partial\beta^T} & \frac{\partial l_\infty}{\partial\beta}\frac{\partial l_\infty}{\partial dH(s)} \\ \frac{\partial l_\infty}{\partial dH(t)}\frac{\partial l_\infty}{\partial\beta^T} & \frac{\partial l_\infty}{\partial dH(t)}\frac{\partial l_\infty}{\partial dH(s)} \end{bmatrix}_{\Omega=\Omega^0},$$

which represents the variance of the score process $n^{1/2}U(\Omega^0)$. In addition, Andersen et al. (2012) showed that for a differentiable functional $F(\Omega)$, by functional delta method, $n^{1/2}\{F(\hat{\Omega})-F(\Omega)\}$ converges weakly to a zero-mean Gaussian process with variance-covariance function $\dot{F}(\Omega)^T I_\infty^{-1}\dot{F}(\Omega)$, where $\dot{F}(\Omega)=\frac{\partial F}{\partial\Omega}$. Applying it to the linear functional defined in Theorem II.2, and replacing I_∞ by its consistent estimator $\mathcal{I}_n$, we can have the stated results.

A.5.4 Proof of Theorem II.3

As we defined in the proof of Theorem II.2,

$$\mathcal{I}_\infty(t,s)=-\begin{bmatrix} \frac{\partial^2 l_\infty}{\partial\beta\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)} \\ \frac{\partial^2 l_\infty}{\partial dH(t)\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)} \end{bmatrix}_{\Omega=\Omega^0}=\begin{bmatrix} \mathcal{I}_{\beta\beta} & \mathcal{I}_{\beta H} \\ \mathcal{I}_{H\beta} & \mathcal{I}_{HH} \end{bmatrix},$$

is the asymptotic covariance matrix operator of the score for the full model. Apply the formula of block matrix inverse to the matrix operator, then we have the inverse of $\mathcal{I}_\infty$ as

$$\mathcal{I}_\infty^{-1}=\begin{bmatrix} Q^{-1} & -Q^{-1}\mathcal{I}_{H\beta}\mathcal{I}_{HH}^{-1} \\ -\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}Q^{-1} & \mathcal{I}_{HH}^{-1}+\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}Q^{-1}\mathcal{I}_{\beta H}\mathcal{I}_{HH}^{-1} \end{bmatrix},$$

where $Q = \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta H}\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}$.

In this part, we need to justify $I_{\beta\beta}^{pr} = \left(- \frac{\partial^2 l_{pr,n}}{\partial\beta\partial\beta^T}\Big|_{\beta=\hat{\beta}} \right)^{-1} \xrightarrow{p} \text{var}\left[\sqrt{n}(\hat{\beta} - \beta^0)\right]$, where $l_{pr,n} = n^{-1}\sum_{i=1}^{n} l_{pr,i}(\beta) = n^{-1}\sum_{i=1}^{n} l_i\big(\beta, \{d\hat{H}(\beta)\}\big)$. We prove it in two steps: First, we show that Q^{-1},the $\beta\beta$ submatrix of $\mathcal{I}_{\infty}^{-1}$, is the limiting covariance matrix of $\hat{\beta}$, i.e., var $\left[\sqrt{n}(\hat{\beta} - \beta^0)\right] = Q^{-1}$. Then, we prove that $I_{\beta\beta}^{pr\,-1} \xrightarrow{p} Q^{-1}$.

First, the proof of var $\left[\sqrt{n}(\hat{\beta} - \beta^0)\right] = Q^{-1}$ is as follows:

Based on Appendix A.3, we have

$$U_{\Omega}(\Omega^0) = \frac{\partial l_n}{\partial\Omega}\Big|_{\Omega=\Omega^0} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\partial \log d\Lambda_i(t)}{\partial\Omega} dM_i(t),$$

which is a martingale under the true model.

The predictable variation process for $\sqrt{n}U_{\Omega}(\Omega^0)$ is

$$< \sqrt{n}U_{\Omega}(\Omega^0) >= \widehat{\text{cov}}[\sqrt{n}U_{\Omega}(\Omega^0)] = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\partial \log d\Lambda_i(t)}{\partial\Omega}\frac{\partial \log d\Lambda_i(t)}{\partial\Omega^T} Y_i(t) d\Lambda_i(t). \tag{A.5}$$

The negative Hessian of the log-likelihood with respect to Ω can be written as

$$\begin{aligned}
&\mathcal{I}_{\Omega\Omega^T}(\Omega^0) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \left\{ \frac{\partial^2 \log d\Lambda_i(t)}{\partial\Omega\partial\Omega^T}\big[dN_i(t) - Y_i(t)d\Lambda_i(t)\big] - \frac{\partial \log d\Lambda_i(t)}{\partial\Omega} Y_i(t) \frac{\partial d\Lambda_i(t)}{\partial\Omega} \right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \left\{ \frac{\partial \log d\Lambda_i(t)}{\partial\Omega}\frac{\partial \log d\Lambda_i(t)}{\partial\Omega^T} Y_i(t) d\Lambda_i(t) - \frac{\partial^2 \log d\Lambda_i(t)}{\partial\Omega\partial\Omega^T} dM_i(t) \right\},
\end{aligned}$$

where the second term turns into an $o_p(1)$, and the first term is a consistent estimator of cov$[\sqrt{n}U_{\Omega}(\Omega^0)]$, as shown in equation (A.5). Therefore, we have

$$\text{cov}[\sqrt{n}U(\Omega^0)] = \mathcal{I}_{\infty} + o_p(1).$$

Since

$$\sqrt{n}(\hat{\Omega} - \Omega^0) = \mathcal{I}_\infty^{-1}\sqrt{n}U(\Omega^0) + o_p(1),$$

then

$$\text{var}[\sqrt{n}(\hat{\Omega} - \Omega^0)] = \mathcal{I}_\infty^{-1}\,\text{cov}[\sqrt{n}U(\Omega^0)]\mathcal{I}_\infty^{-1} + o_p(1) = \mathcal{I}_\infty^{-1}.$$

Thus,

$$\text{var}[\sqrt{n}(\hat{\beta} - \beta^0)] = Q^{-1}.$$

Next, we prove that $I_{\beta\beta}^{pr\,-1} \xrightarrow{p} Q^{-1}$:

Denote the Jacobian $\dfrac{\partial d\hat{H}_\beta(s)}{\partial \beta}$ as $J_{H\beta}$. The profile score function for β is

$$U_\beta^{pr} = \frac{dl_{pr}}{d\beta} = \frac{dl\big(\beta, \{d\hat{H}_\beta(s)\}\big)}{d\beta} = \left(\frac{\partial l}{\partial \beta} + \int_s \frac{\partial l}{\partial dH(s)}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\right)\Bigg|_{dH=d\hat{H}_\beta} = \frac{\partial l}{\partial \beta}\Bigg|_{dH=d\hat{H}_\beta}$$

Since $d\hat{H}_\beta$ solves the score equation $\dfrac{\partial l}{\partial dH(s)} = 0$, we have $\dfrac{\partial l}{\partial dH(s)}\Big|_{dH=d\hat{H}_\beta} = 0.$

The negative Hessian of the profile log-likelihood can be derived as

$$
\begin{aligned}
I^{pr}_{\beta\beta} =& - \frac{d^2 l_{pr}}{d\beta d\beta^T}\Big|_{dH=d\hat{H}_\beta} \\
=& - \frac{\partial}{\partial dH(t)}\left[\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\right]\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial}{\partial \beta^T}\left[\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\right] \\
& - \frac{\partial}{\partial dH(t)}\frac{\partial l}{\partial \beta}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big|_{d\hat{H}_\beta} \\
=& - \frac{\partial^2 l}{\partial dH(s)\partial dH(t)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial^2 d\hat{H}_\beta(s)}{\partial\beta\partial dH(t)}\frac{\partial d\hat{H}_\beta(t)}{\partial\beta^T} \\
& - \frac{\partial^2 l}{\partial dH(s)\partial\beta^T}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial\beta} - \frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial^2 d\hat{H}_\beta(s)}{\partial\beta\partial\beta^T} - \frac{\partial^2 l}{\partial\beta\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial\beta^T} \\
& - \frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big|_{d\hat{H}_\beta} \\
=& J_{\beta H}\hat{I}_{HH}J_{H\beta} - 0 + J_{\beta H}\hat{I}_{H\beta} - 0 + \hat{I}_{\beta H}J_{H\beta} + \hat{I}_{\beta\beta} \\
=& J_{\beta H}\hat{I}_{HH}J_{H\beta} + J_{\beta H}\hat{I}_{H\beta} + \hat{I}_{\beta H}J_{H\beta} + \hat{I}_{\beta\beta}.
\end{aligned}
$$

To express $I^{pr}_{\beta\beta}$ without the Jacobians, we make use of the fact that $\dfrac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta} = 0$, such that

$$
\begin{aligned}
& \frac{d}{d\beta}\left(\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\right) = 0 \\
=& \frac{\partial^2 l}{\partial dH(s)\partial\beta}\Big|_{d\hat{H}_\beta} + \frac{\partial^2 l}{\partial dH(s)\partial dH(t)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(t)}{\partial\beta} = -\hat{I}_{H\beta} - \hat{I}_{HH}J_{H\beta} \\
\Longrightarrow\quad & J_{H\beta} = -\hat{I}^{-1}_{HH}\hat{I}^{-1}_{H\beta}
\end{aligned}
$$

Replace the Jacobians in $I^{pr}_{\beta\beta}$ with the expression above, we have

$$
I^{pr}_{\beta\beta} = \hat{I}_{\beta\beta} - \hat{I}_{\beta H}\hat{I}^{-1}_{HH}\hat{I}_{H\beta} \quad \xrightarrow{p} \quad \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta H}\mathcal{I}^{-1}_{HH}\mathcal{I}_{H\beta} = Q.
$$

Thus,

$$
I^{pr\,-1}_{\beta\beta} \xrightarrow{p} Q^{-1} = \mathrm{var}[\sqrt{n}(\hat{\beta} - \beta^0)].
$$

APPENDIX B

A Semiparametric Joint Model for Estimating the Screening Effect on Cancer Mortality

B.1 Prediction of survival function for the terminal event

This section presents the derivation of conditional survival functions of terminal event given incidence information at time t^*. Specifically, we are interested in $Pr(T_D > t|X_1 = t^*, \Delta_1)$. For $t \geq t^*$,

1. Consider a subject that has been diagnosed at t^*, i.e. $\Delta_1 = 1, T_I = t^*$

$$Pr(T_D > t|T_I = t^*) = \frac{Pr(T_I = t^*, T_D > t)}{Pr(T_I = t^*)} = \frac{L_{10}(t^*, t)}{L_{10}(t^*, t^*)}$$

$$= \frac{\int_t^{t_{lf2}} f_{T_D}(t_D) f_{T_I|T_D}(t^*|t_D)\, dt_D + f_{T_D}(\infty) f_{T_I|T_D}(t^*|\infty)}{\int_{t^*}^{t_{lf2}} f_{T_D}(t_D) f_{T_I|T_D}(t^*|t_D)\, dt_D + f_{T_D}(\infty) f_{T_I|T_D}(t^*|\infty)}$$

$$= \frac{\displaystyle\int_t^{t_{lf2}} \frac{\eta e^{-\eta H(t_D)}}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + \frac{e^{-\eta H(t_{lf2})}}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\displaystyle\int_{t^*}^{t_{lf2}} \frac{\eta e^{-\eta H(t_D)}}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\, dH(t_D) + \frac{e^{-\eta H(t_{lf2})}}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}}$$

2. Consider a subject that has not been diagnosed at t^*, i.e. $\Delta_1 = 0, T_I > t^*$

$$Pr(T_D > t | T_I > t^*) = \frac{Pr(T_I > t^*, T_D > t)}{Pr(T_I > t^*)} = \frac{L_{00}(t^*, t)}{L_{00}(t^*, t^*)}$$

$$= \frac{\int_t^{t_{lf2}} \int_{t^*}^{t_D} f_{T_D}(t_D) f_{T_I|T_D}(t_I|t_D)\, dt_I dt_D + \int_{t^*}^{\infty} f_{T_D}(\infty) f_{T_I|T_D}(t_I|\infty)\, dt_I}{\int_{t^*}^{t_{lf2}} \int_{t^*}^{t_D} f_{T_D}(t_D) f_{T_I|T_D}(t_I|t_D)\, dt_I dt_D + \int_{t^*}^{\infty} f_{T_D}(\infty) f_{T_I|T_D}(t_I|\infty)\, dt_I}$$

$$= \frac{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:t^*<t_k\le t_D} h(t_k)\mu(t_k)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>t^*} h(t_k)\mu(t_k)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\int_{t^*}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:t^*<t_k\le t_D} h(t_k)\mu(t_k)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>t^*} h(t_k)\mu(t_k)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}$$

B.2 Derivation of hazard terms

For ease of notation, we suppress the subscript i for each individual subject.

1. The hazard of having incidence at time t:

$$d\Lambda_1(t) = Pr\big(dN_1(t) = 1 | Y_1(t) = 1\big) = \frac{Pr(T_I = t)}{Pr(T_I \ge t)} = \frac{L_{10}(t,t)}{L_{00}(t,t)}$$

$$= \frac{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\mu(t)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\mu(t)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}{\sum_{k:t<t_k\le t_D} h(t_k)\mu(t_k)} h(t)$$
$$\frac{}{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:t<t_k\le t_D} h(t_k)\mu(t_k)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k>t} h(t_k)\mu(t_k)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}}$$

2. The hazard of dying from cancer at time t, given incidence at time t_1:

$$d\Lambda_2(t|t_1) = Pr[dN_2(t|t_1) = 1 | Y_2(t|t_1) = 1]$$

$$= \frac{Pr(T_I = t_1, T_D = t)}{Pr(T_I = t_1, T_D \ge t)} = \frac{L_{11}(t_1, t)}{L_{10}(t_1, t)}$$

$$= \frac{\eta e^{-\eta H(t)} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:0<t_k\le t} h(t_k)\mu(t_k)}}{\int_t^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}\, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{h(t_1)\mu(t_1)}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)}} dH(t)$$

B.3 EM algorithm

This section presents the derivation of EM algorithm to estimate the nonparametric components $\{dH\}$ and $\{h\}$ for our model. When the terminal event is censored, we introduce an artificial variable U to transform the likelihood, and treating this artificial variable and the censored terminal event as missing data, an EM algorithm is derived to obtain $\{dH\}$ and $\{h\}$.

Conditional score function for $dH(s)$ can be written as

$$U_0(s) = \frac{\partial \log L_{com}}{\partial dH(s)} = \frac{\partial \log(L_0 P_0)}{\partial dH(s)}.$$

Here, L_0 is the conditional likelihood of the incidence event, given the terminal event observed at $T_D = t_D$ and U; and P_0 is the joint likelihood of T_D and U.

In the (k+1)th iteration,

E-step: Unconditional score function

$$U(s) = E[U_0(s)||L_0^{(k)}] = \frac{E[U_0(s)L_0^{(k)}]}{E[L_0^{(k)}]}.$$

M-step: Set $U(s) = 0$, then we can obtain $dH^{(k+1)}(s)$ to update the expression of $dH(s)$.

Similar steps are taken to iteratively update the estimates of $h(s)$.

The rest of this section is organized as follows. First, we write out the conditional score function for $\{dH\}$ and $\{h\}$, for the three different cases respectively (both events observed, incidence observed yet death censored, and neither events observed). Next we derive the E step for $\{dH\}$ and $\{h\}$. Finally we derive the M step to maximize the likelihood with respect to $\{dH\}$ and $\{h\}$, which has a closed-form expression analogous to the weighted Breslow-type estimators of Chen (2009).

B.3.1 EM algorithm for $\{dH\}$

B.3.1.1 Conditional score function

Conditional score function w.r.t $\{dH\}$ is derived as

$$U_{0,dH}(s) = \frac{\partial log L_{com}}{\partial dH(s)}.$$

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$):

   Since there is no missingness for observed events, we do not need to introduce an artificial variable,

   $$L_{com} = \eta dH(X_2) e^{-\eta H(X_2)} \frac{h(X_1)\mu(X_1)}{1 + \sum\limits_{k:0<t_k\leq X_2} h(t_k)\mu(t_k)}.$$

   Then we have

   $$U_{0,dH(s)} = \frac{\mathbb{1}(X_2 = s)}{dH(s)} - \eta \mathbb{1}(X_2 \geq s).$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 0$):

   (a) If $X_2 \geq t_{lf2}$:

   In this case, terminal event occurs at time infinity, and there is no missing data,

   $$L_{com} = e^{-\eta H(t_{lf2})} \frac{h(X_1)\mu(X_1)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}.$$

   Then,

   $$U_{0,dH(s)} = -\eta \mathbb{1}(t_{lf2} \geq s).$$

   (b) If $X_2 < t_{lf2}$:

   Since terminal event T_D is not observed, an artificial variable U is introduced to simplify computation, and the complete data is ($T_I = X_1, T_D =$

$t_D, U = u)$,

$$L_{com} = \begin{cases} h(X_1)\mu(X_1)\eta dH(t_D)\exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}, & t_D \leq t_{lf2} \\ h(X_1)\mu(X_1)\exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]\}, & t_D = \infty. \end{cases}$$

Then we can have

$$U_{0,dH(s)} = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH(s)} - \eta\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ -\eta\mathbb{1}(t_{lf2} \geq s), & T_D = \infty. \end{cases}$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = 0$, $\Delta_2 = 0$):

   (a) If $X_2 \geq t_{lf1}$:

   In this case, both incidence and death occur at time infinity, and

   $$L_{com} = e^{-\eta H(t_{lf2})} \frac{h(\infty)\mu(\infty)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)}.$$

   Then,

   $$U_{0,dH(s)} = -\eta\mathbb{1}(t_{lf2} \geq s).$$

   (b) If $X_2 < t_{lf1}$ and $X_2 \geq t_{lf2}$:

   In this case, incidence event T_I occurs after time X_2, and death T_D occurs at time infinity, we have the complete data likelihood as

   $$L_{com} = \frac{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)} e^{-\eta H(t_{lf2})}.$$

   Then,

   $$U_{0,dH(s)} = -\eta\mathbb{1}(t_{lf2} \geq s).$$

(c) If $X_2 < t_{lf1}$ and $X_2 < t_{lf2}$:

Incidence and death events are both censored, and likelihood for the complete data $(T_I > X_2, T_D, U)$ is

$$L_{com} =$$

$$\begin{cases} [\sum_{k:X_2<t_k\le t_D} h(t_k)\mu(t_k)]\eta dH(t_D)\exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)]\}, & t_D \le t_{lf2} \\ [\sum_{k:t_k>X_2} h(t_k)\mu(t_k)]\exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]\}, & t_D = \infty. \end{cases}$$

Then, we have

$$U_{0,dH(s)} = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH(s)} - \eta\mathbb{1}(T_D \ge s), & T_D \le t_{lf2} \\ -\eta\mathbb{1}(t_{lf2} \ge s), & T_D = \infty. \end{cases}$$

B.3.1.2 E step

The unconditional score function for $\{dH\}$ is

$$U_{dH}(s) = E[U_{0,dH}(s)||L_0^{(k)}] = \frac{E[U_{0,dH}(s)L_0^{(k)}]}{E[L_0^{(k)}]}.$$

In the rest of EM algorithm section, we keep iteration index for $(k+1)^{th}$ iteration when needed, and drop index for k^{th} for brevity. Note that the denominator $E[L_0^{(k)}]$ is just observed data likelihood, as in Section 3.2.4.

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$)

   Since there is no missingness, the unconditional score function is

$$U_{dH(s)} = U_{0,dH(s)} = \frac{dN_2(s)}{dH(s)} - \eta Y_2(s). \tag{B.1}$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_=1$, $\Delta_2 = 0$)

(a) If $X_2 \geq t_{lf2}$:

$$U_{dH(s)} = U_{0,dH(s)} = -\eta \mathbb{1}(t_{lf2} \geq s).$$

(b) If $X_2 < t_{lf2}$:

First, we write out L_0 and P_0. P_0 is the distribution of missing data (T_D, U):

$$P_0 = f_{T_D}(t_D) f_U(u) = \begin{cases} \eta dH(t_D) e^{-\eta H(t_D)-u}, & t_D \leq t_{lf2} \\ e^{-\eta H(t_{lf2})-u}, & t_D = \infty. \end{cases}$$

L_0 is the conditional incidence given T_D and U, and in this case,

$$L_0 = f_{T_I|T_D,U}(X_1|t_D, u) = h(X_1)\mu(X_1) \exp\{-u \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)\}.$$

Next, we calculate the numerator $E[U_{0,dH}(s)L_0]$, and finally obtain the unconditional score function $U_{dH}(s)$. Note, the denominator $E[L_0^{(k)}]$ is the observed data likelihood

$$\begin{aligned} & E[L_0] = L_{obs,10} \\ = & \int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{h(X_1)\mu(X_1)}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)} \, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{h(X_1)\mu(X_1)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)} \\ & = h(X_1)\mu(X_1)[U(X_2, t_{lf2}) + V]. \end{aligned}$$

We consider the calculation in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf2}$, and $s > t_{lf2}$, respectively.

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  First, based on conditional score function derived in Appendix B.3.1.1, $U_{0,dH(s)} = -\eta$, which does not depend on any missing data.

Thus we can have

$$U_{dH_I}(s) = \frac{E[U_{0,dH_I}(s)L_0]}{E[L_0]} = U_{0,dH_I}(s)\frac{E[L_0]}{E[L_0]} = -\eta.$$

- $X_2 < s \leq t_{lf2}$, i.e. $Y_2(s) = 0$, $Y_{lf2}(s) = 1$

  First,

$$U_{0,dH(s)} = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH(s)} - \eta\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ -\eta, & T_D = \infty. \end{cases}$$

  Then the numerator $E[U_{0,dH}(s)L_0]$

$$= \frac{E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \leq t_{lf2})L_0]}{dH^{(k+1)}(s)} - \eta E[\mathbb{1}(s \leq T_D \leq t_{lf2})L_0] - \eta E[\mathbb{1}(T_D = \infty)L_0]$$

We can derive the term $E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \leq t_{lf2})L_0]$ as

$$\begin{aligned} &E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \leq t_{lf2})L_0] \\ &= \int_0^\infty h(X_1)\mu(X_1)\eta dH(s)\exp\{-\eta H(s) - u[1 + \sum_{k:0<t_k\leq s} h(t_k)\mu(t_k)]\}du \\ &= \eta dH(s)e^{-\eta H(s)}\frac{h(X_1)\mu(X_1)}{1 + \sum_{k:0<t_k\leq s} h(t_k)\mu(t_k)} = h(X_1)\mu(X_1)U.sdH(s). \end{aligned}$$

The term $E[\mathbb{1}(s \leq T_D \leq t_{lf2})L_0]$ can be derived as

$$\begin{aligned} &E[\mathbb{1}(s \leq T_D \leq t_{lf2})L_0] \\ &= \int_0^\infty \int_{s^-}^{t_{lf2}} h(X_1)\mu(X_1)\eta\exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}dH(t_D)du \\ &= \int_{s^-}^{t_{lf2}} \eta e^{-\eta H(t_D)}\frac{h(X_1)\mu(X_1)}{1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}dH(t_D) = h(X_1)\mu(X_1)U(s^-, t_{lf2}). \end{aligned}$$

The term $E[\mathbb{1}(T_D = \infty)L_0]$ is

$$\begin{aligned}
&E[\mathbb{1}(T_D = \infty)L_0] \\
=&\int_0^\infty h(X_1)\mu(X_1)\exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]\}du \\
=&e^{-\eta H(t_{lf2})}\frac{h(X_1)\mu(X_1)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)} = h(X_1)\mu(X_1)V.
\end{aligned}$$

Then the numerator can be calculated as

$$\begin{aligned}
E[U_{0,dH}(s)L_0] =& \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]h(X_1)\mu(X_1)U.s \\
&- h(X_1)\mu(X_1)[\eta U(s^-, t_{lf2}) + \eta V - U.s].
\end{aligned}$$

Thus, the unconditional score function for $dH(s)$ is

$$U_{dH((s)} = \frac{\left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]U.s - [\eta U(s^-, t_{lf2}) + \eta V - U.s]}{U(X_2, t_{lf2}) + V}.$$

- $s > t_{lf2}$, i.e. $Y_{lf2}(s) = 0$

  Since $U_{0,dH(s)} = 0$, the unconditional score function $U_{dH(s)}$ is also zero.

Combining these results, the contribution of a subject, with intermediate event observed at X_1 yet terminal event censored at X_2, to the unconditional score function, can be written as

$$\begin{aligned}
U_{dH}(s) =& \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]\mathbb{1}(X_2 < t_{lf2})[1 - Y_2(s)]\frac{U.s}{U(X_2, t_{lf2}) + V} \\
&-\left\{\mathbb{1}(X_2 \geq t_{lf2})\eta + \mathbb{1}(X_2 < t_{lf2})\left\{Y_2(s)\eta + [1 - Y_2(s)]\frac{\eta[U(s^-, t_{lf2}) + V] - U.s}{U(X_2, t_{lf2}) + V}\right\}\right\}.
\end{aligned} \tag{B.2}$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = 0$, $\Delta_2 = 0$)

(a) If $X_2 \geq t_{lf1}$:

$$U_{dH(s)} = U_{0,dH(s)} = -\eta \mathbb{1}(t_{lf2} \geq s).$$

(b) If $X_2 < t_{lf1}$ and $X_2 \geq t_{lf2}$:

$$U_{dH(s)} = U_{0,dH(s)} = -\eta \mathbb{1}(t_{lf2} \geq s).$$

(c) If $X_2 < t_{lf1}$ and $X_2 < t_{lf2}$:

First we need P_0 and L_0. P_0 keeps the same distribution as before, and

$$L_0 = Pr(T_I > X_2 | t_D, u) = [\sum_{k:X_2 < t_k \leq t_D} h(t_k)\mu(t_k)] \exp\{-u \sum_{k:0 < t_k \leq t_D} h(t_k)\mu(t_k)\}.$$

Then we calculate $U_{dH(s)}$ in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf2}$, and $s > t_{lf2}$. Note, the denominator $E[L_0^{(k)}]$ is the observed data likelihood

$$\begin{aligned} & E[L_0] = L_{obs,00} \\ &= \int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum_{k:X_2 < t_k \leq t_D} h(t_k)\mu(t_k)}{1 + \sum_{k:0 < t_k \leq t_D} h(t_k)\mu(t_k)} \, dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum_{k:t_k > X_2} h(t_k)\mu(t_k)}{1 + \sum_{k:t_k > 0} h(t_k)\mu(t_k)} \\ & \qquad = W(X_2, t_{lf2}) + Z. \end{aligned}$$

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  Since $U_{0,dH(s)} = -\eta$, $U_{dH(s)} = -\eta$.

- $X_2 < s \leq t_{lf2}$, i.e. $Y_2(s) = 0$, $Y_{lf2}(s) = 1$

  To calculate $U_{dH(s)}$, first

$$U_{0,dH(s)} = \begin{cases} \dfrac{\mathbb{1}(T_D = s)}{dH(s)} - \eta \mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ -\eta, & T_D = \infty. \end{cases}$$

Then the numerator $E[U_{0,dH}(s)L_0]$ is

$$\frac{E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \le t_{lf2})L_0]}{dH^{(k+1)}(s)} - \eta E[\mathbb{1}(s \le T_D \le t_{lf2})L_0] - \eta E[\mathbb{1}(T_D = \infty)L_0].$$

We can derive the term $E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \le t_{lf2})L_0]$ as

$$\begin{aligned}
&E[\mathbb{1}(T_D = s)\mathbb{1}(T_D \le t_{lf2})L_0] \\
&= \int_0^\infty [\sum_{k:X_2<t_k\le s} h(t_k)\mu(t_k)]\eta dH(s)\exp\{-\eta H(s) - u[1 + \sum_{k:0<t_k\le s} h(t_k)\mu(t_k)]\}du \\
&= \eta dH(s)e^{-\eta H(s)}\frac{\sum_{k:X_2<t_k\le s} h(t_k)\mu(t_k)}{1 + \sum_{k:0<t_k\le s} h(t_k)\mu(t_k)} = W.sdH(s).
\end{aligned}$$

The term $E[\mathbb{1}(s \le T_D \le t_{lf2})L_0]$ can be derived as

$$\begin{aligned}
&E[\mathbb{1}(s \le T_D \le t_{lf2})L_0] = \\
&\int_0^\infty \int_{s^-}^{t_{lf2}} [\sum_{k:X_2<t_k\le t_D} h(t_k)\mu(t_k)]\eta\exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)]\}dH(t_D)du \\
&= \int_{s^-}^{t_{lf2}} \eta e^{-\eta H(t_D)}\frac{\sum_{k:X_2<t_k\le t_D} h(t_k)\mu(t_k)}{1 + \sum_{k:0<t_k\le t_D} h(t_k)\mu(t_k)}dH(t_D) = W(s^-, t_{lf2}).
\end{aligned}$$

The term $E[\mathbb{1}(T_D = \infty)L_0]$ is

$$\begin{aligned}
&E[\mathbb{1}(T_D = \infty)L_0] \\
&= \int_0^\infty [\sum_{k:t_k>X_2} h(t_k)\mu(t_k)]\exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]\}du \\
&= e^{-\eta H(t_{lf2})}\frac{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)}{1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)} = Z.
\end{aligned}$$

Then the numerator can be calculated as

$$E[U_{0,dH}(s)L_0] = \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]W.s - [\eta W(s^-, t_{lf2}) + \eta Z - W.s].$$

Thus, the unconditional score function for $dH(s)$ is

$$U_{dH((s)} = \frac{\left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]W.s - [\eta W(s^-, t_{lf2}) + \eta Z - W.s]}{W(X_2, t_{lf2}) + Z}.$$

- $s > t_{lf2}$, i.e. $Y_{lf2}(s) = 0$

  Since $U_{0,dH(s)} = 0$, $U_{dH(s)}$ is also zero.

Combining these results, the contribution of a subject, with both intermediate and terminal events censored at X_2, to the unconditional score function, can be written as

$$\begin{aligned} U_{dH}(s) =& \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]\mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})[1 - Y_2(s)]\frac{W.s}{W(X_2, t_{lf2}) + Z} \\ &- \Bigg\{\mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\Big\{Y_2(s)\eta + [1 - Y_2(s)]\frac{\eta[W(s^-, t_{lf2}) + Z] - W.s}{W(X_2, t_{lf2}) + Z}\Big\} \\ &+ \Big[1 - \mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\Big]\eta\Bigg\}. \end{aligned} \tag{B.3}$$

Combining equations (B.1), (B.2) and (B.3), we have unconditional score function for $dH(s)$ as

$$U_{dH^{(k+1)}}(s) = \frac{dN_2(s)}{dH^{(k+1)}(s)} - \Psi_{dH}^{(k)}(s) + \left[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\right]\theta_{dH}^{(k)}(s),$$

where

$$\begin{aligned}
\Psi_{dH}(s) =& \Delta_2 Y_2(s)\eta \\
&+\Delta_1(1-\Delta_2)\mathbb{1}(X_2 < t_{lf2})\Big\{ -\eta + Y_2(s)\eta + [1-Y_2(s)]\frac{\eta[U(s^-, t_{lf2})+V] - U.s}{U(X_2, t_{lf2})+V}\Big\} \\
&+(1-\Delta_1)\mathbb{1}(X_1 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\Big\{ -\eta + Y_2(s)\eta \\
&\qquad + [1-Y_2(s)]\frac{\eta[W(s^-, t_{lf2})+Z] - W.s}{W(X_2, t_{lf2})+Z}\Big\}, \\
\theta_{dH}(s) =& \mathbb{1}(X_2 < t_{lf2})[1-Y_2(s)](1-\Delta_2)\Big\{\Delta_1 \frac{U.s}{U(X_2, t_{lf2})+V} \\
&\qquad + \mathbb{1}(X_1 < t_{lf1})(1-\Delta_1)\frac{W.s}{W(X_2, t_{lf2})+Z}\Big\}.
\end{aligned}$$

B.3.1.3 M step

Suppose we have n independent subjects with observed data $(X_{1i}, \Delta_{1i}, X_{2i}, \Delta_{2i})$, $i = 1, 2, ..., n$. The estimator of $dH^{(k+1)}(s)$ can be derived by solving $\sum_{i=1}^{n} U_{i,dH}(s) = 0$. Now we want to solve

$$\sum_{i=1}^{n}\left\{\frac{dN_{2i}(s)}{dH^{(k+1)}(s)} - \Psi_{dH,i}(s) + \Big[\frac{dH^{(k)}(s)}{dH^{(k+1)}(s)} - 1\Big]\theta_{dH,i}(s)\right\} = 0,$$

which gives

$$dH^{(k+1)}(s) = \frac{\sum_i dN_{2i}(s) + \Big[\sum_i \theta^{(k)}_{dH,i}(s)\Big]dH^{(k)}(s)}{\sum_i [\Psi^{(k)}_{dH,i}(s) + \theta^{(k)}_{dH,i}(s)]}.$$

The above equation solves for $dH(s)$ iteratively until convergence, and the estimator at convergence, $\widehat{dH}(s)$, is a consistent estimator of $dH(s)$ (Tsodikov (2003)).

B.3.2 EM algorithm for $\{h\}$

B.3.2.1 Conditional score function

Conditional score function w.r.t $\{h\}$ is derived as

$$U_{0,h}(s) = \frac{\partial log L_{com}}{\partial h(s)}.$$

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$):

$$U_{0,h(s)} = \frac{\mathbb{1}(X_1 = s)}{h(s)} - \frac{\mu(s)}{1 + \sum\limits_{k:0<t_k \leq X_2} h(t_k)\mu(t_k)} \mathbb{1}(X_2 \geq s).$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 0$):

   (a) If $X_2 \geq t_{lf2}$:

$$U_{0,h(s)} = \frac{\mathbb{1}(X_1 = s)}{h(s)} - \frac{\mu(s)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}.$$

   (b) If $X_2 < t_{lf2}$:

$$U_{0,h(s)} = \begin{cases} \dfrac{\mathbb{1}(X_1 = s)}{h(s)} - \mu(s)U\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ \dfrac{\mathbb{1}(X_1 = s)}{h(s)} - \mu(s)U, & T_D = \infty. \end{cases}$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = 0$, $\Delta_2 = 0$):

   (a) If $X_2 \geq t_{lf1}$:

$$U_{0,h(s)} = \frac{\mathbb{1}(s = \infty)}{h(\infty)} - \frac{\mu(s)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}.$$

(b) If $X_2 < t_{lf1}$ and $X_2 \geq t_{lf2}$:

$$U_{0,h(s)} = \frac{\mu(s)}{\sum\limits_{k:t_k > X_2} h(t_k)\mu(t_k)} \mathbb{1}(X_2 < s) - \frac{\mu(s)}{1 + \sum\limits_{k:t_k > 0} h(t_k)\mu(t_k)}.$$

(c) If $X_2 < t_{lf1}$ and $X_2 < t_{lf2}$:

$$U_{0,h(s)} = \begin{cases} \dfrac{\mathbb{1}(X_2 < s \leq T_D)}{\sum\limits_{k:X_2 < t_k \leq T_D} h(t_k)\mu(t_k)} - \mu(s)U\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ \dfrac{\mu(s)}{\sum\limits_{k:t_k > X_2} h(t_k)\mu(t_k)} \mathbb{1}(X_2 < s) - \mu(s)U, & T_D = \infty. \end{cases}$$

B.3.2.2 E step

The unconditional score function for $\{h\}$ is

$$U_h(s) = E[U_{0,h}(s)||L_0^{(k)}] = \frac{E[U_{0,h}(s)L_0^{(k)}]}{E[L_0^{(k)}]}.$$

1. Subject has incidence at X_1, and dies at X_2 (i.e. $\Delta_1 = 1$, $\Delta_2 = 1$)

   Since there is no missingness, the unconditional score function is

$$U_{h(s)} = U_{0,h(s)} = \frac{dN_1(s)}{h(s)} - \frac{\mu(s)}{1 + \sum\limits_{k:0 < t_k \leq X_2} h(t_k)\mu(t_k)} Y_2(s). \tag{B.4}$$

2. Subject has incidence at X_1, and is censored at X_2 (i.e. $\Delta_I = 1$, $\Delta_D = 0$)

   (a) If $X_2 \geq t_{lf2}$:

$$U_{h(s)} = U_{0,h(s)} = \frac{dN_1(s)}{h(s)} - \frac{\mu(s)}{1 + \sum\limits_{k:t_k > 0} h(t_k)\mu(t_k)}.$$

   (b) If $X_2 < t_{lf2}$:

   First we have the conditional score function $U_{0,h(s)}$as in Appendix B.3.1.1.

Next, we calculate the numerator $E[U_{0,dH}(s)L_0]$, and finally obtain the unconditional score function $U_{dH}(s)$. We consider the calculation in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf2}$, and $s > t_{lf2}$, respectively.

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  First,

  $$U_{0,h(s)} = \frac{dN_1(s)}{h(s)} - \mu(s)U$$

  Then the numerator $E[U_{0,h}(s)L_0]$ is

  $$E[U_{0,h}(s)L_0] = \frac{dN_1(s)}{h(s)}E[L_0] - \mu(s)E[UL_0].$$

  The term $E[UL_0]$ can be derived as

$$\begin{aligned}
E[UL_0] = &\int_0^{\infty}\int_{X_2}^{t_{lf2}} uh(X_1)\mu(X_1)\eta exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}dH(t_D)du \\
&+ \int_0^{\infty} uh(X_1)\mu(X_1)exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]\}du \\
= &\int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{h(X_1)\mu(X_1)}{[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]^2} dH(t_D) + e^{-\eta H(t_{lf2})} \frac{h(X_1)\mu(X_1)}{[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]^2} \\
&= h(X_1)\mu(X_1)[P(X_2, t_{lf2}) + Q].
\end{aligned}$$

  Then the numerator can be calculated as

  $$E[U_{0,h}(s)L_0] = \frac{dN_1(s)}{h(s)}E[L_0] - \mu(s)h(X_1)\mu(X_1)[P(X_2, t_{lf2}) + Q].$$

  Thus, the unconditional score function for $h(s)$ is

  $$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = \frac{dN_1(s)}{h(s)} - \mu(s)\frac{P(X_2, t_{lf2}) + Q}{U(X_2, t_{lf2}) + V}.$$

- $X_2 < s \leq t_{lf2}$, i.e. $Y_2(s) = 0$, $Y_{lf2}(s) = 1$

  First, the conditional score function for $h(s)$

$$U_{0,h(s)} = \begin{cases} -\mu(s)U\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ -\mu(s)U, & T_D = \infty. \end{cases}$$

  Then the numerator $E[U_{0,h(s)}L_0]$ can be written as

$$E[U_{0,h(s)}L_0] = -\mu(s)E[\mathbb{1}(s \leq T_D \leq t_{lf2})UL_0] - \mu(s)E[\mathbb{1}(T_D = \infty)UL_0].$$

  We can derive the term $E[\mathbb{1}(s \leq T_D \leq t_{lf2})UL_0]$ as

$$\int_0^\infty \int_{s^-}^{t_{lf2}} uh(X_1)\mu(X_1)\eta exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}dH(t_D)du$$
$$= \int_{s^-}^{t_{lf2}} \frac{h(X_1)\mu(X_1)}{[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]^2}\eta e^{-\eta H(t_D)}dH(t_D) = h(X_1)\mu(X_1)P(s^-, t_{lf2}).$$

  The term $E[\mathbb{1}(T_D = \infty)UL_0]$ is

$$E[\mathbb{1}(T_D = \infty)UL_0] = e^{-\eta H(t_{lf2})}\frac{h(X_1)\mu(X_1)}{[1 + \sum_{k:t_k>0} h(t_k)\mu(t_k)]^2} = h(X_1)\mu(X_1)Q.$$

  So the numerator is calculated as

$$E[U_{0,h(s)}L_0] = -\mu(s)h(X_1)\mu(X_1)[P(s^-, t_{lf2}) + Q].$$

  Thus the unconditional score function for $h(s)$ is

$$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = -\mu(s)\frac{P(s^-, t_{lf2}) + Q}{U(X_2, t_{lf2}) + V}.$$

- $s > t_{lf2}$, i.e. $Y_{lf2}(s) = 0$

  The conditional score function for $h(s)$ is

$$U_{0,h(s)} = \begin{cases} 0, & T_D \leq t_{lf2} \\ -\mu(s)U, & T_D = \infty. \end{cases}$$

  Then the numerator $E[U_{0,h(s)}L_0]$ can be calculated as

$$E[U_{0,h(s)}L_0] = -\mu(s)E[\mathbb{1}(T_D = \infty)UL_0] = -\mu(s)h(X_1)\mu(X_1)Q.$$

  Thus the unconditional score function for $h(s)$ is

$$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = -\mu(s)\frac{Q}{U(X_2, t_{lf2}) + V}.$$

Combining these results, the contribution of a subject, with intermediate event observed at X_1 yet terminal event censored at X_2, to the unconditional score function for $h(s)$, can be written as

$$U_{h(s)} = \frac{dN_1(s)}{h^{(k+1)}(s)} - \mu(s)\mathbb{1}(X_2 < t_{lf2})\Bigg\{ - \frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} + \frac{Q + Y_2(s)P(X_2, t_{lf2}) + [1 - Y_2(s)]Y_{lf2}(s)P(s^-, t_{lf2})}{U(X_2, t_{lf2}) + V} \Bigg\}. \quad \text{(B.5)}$$

3. Subject is censored at X_2 before any event is observed (i.e. $\Delta_1 = 0$, $\Delta_2 = 0$)

   (a) If $X_2 \geq t_{lf1}$:

$$U_{h(s)} = U_{0,h(s)} = \frac{\mathbb{1}(s = \infty)}{h(\infty)}\mu(\infty) - \frac{\mu(s)}{1 + \sum\limits_{k:t_k>0} h(t_k)\mu(t_k)}.$$

(b) If $X_2 < t_{lf1}$ and $X_2 \geq t_{lf2}$:

$$U_{h(s)} = U_{0,h(s)} = \frac{\mathbb{1}(X_2 < s)}{\sum\limits_{k:t_k > X_2} h(t_k)\mu(t_k)}\mu(s) - \frac{\mu(s)}{1 + \sum\limits_{k:t_k > 0} h(t_k)\mu(t_k)}.$$

(c) If $X_2 < t_{lf1}$ and $X_2 < t_{lf2}$:

We consider $U_{dH(s)}$ in three cases, $s \leq X_2$, $X_2 < s \leq t_{lf2}$, and $s > t_{lf2}$, respectively. Note, the denominator $E[L_0^{(k)}]$ is the observed data likelihood

$$E[L_0] = L_{obs,00} = W(X_2, t_{lf2}) + Z.$$

- $s \leq X_2$, i.e. $Y_2(s) = 1$

  First, conditional score function is $U_{0,h(s)} = -\mu(s)U$.

  Then the numerator $E[U_{0,h(s)}L_0]$ can be calculated as

$$E[U_{0,h(s)}L_0] = -\mu(s)E[UL_0]$$

$$= -\mu(s)\int_0^{\infty}\int_{X_2}^{t_{lf2}} u[\sum_{k:t_k > X_2} h(t_k)\mu(t_k)]\eta exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k \leq t_D} h(t_k)\mu(t_k)]\}dH(t_D)du$$

$$- \mu(s)\int_0^{\infty} u[\sum_{k:t_k > X_2} h(t_k)\mu(t_k)]exp\{-\eta H(t_{lf2}) - u[1 + \sum_{k:t_k > 0} h(t_k)\mu(t_k)]\}du$$

$$= \int_{X_2}^{t_{lf2}} \eta e^{-\eta H(t_D)} \frac{\sum\limits_{k:t_k > X_2} h(t_k)\mu(t_k)}{[1 + \sum\limits_{k:0<t_k \leq t_D} h(t_k)\mu(t_k)]^2} dH(t_D) + e^{-\eta H(t_{lf2})} \frac{\sum\limits_{k:t_k > X_2} h(t_k)\mu(t_k)}{[1 + \sum\limits_{k:t_k > 0} h(t_k)\mu(t_k)]^2}$$

$$= -\mu(s)(R + V - T).$$

Thus unconditional score function $U_{h(s)}$ is derived as

$$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = -\mu(s)\frac{R + V - T}{W(X_2, t_{lf2}) + Z}.$$

- $X_2 < s \leq t_{lf2}$, i.e. $Y_2(s) = 0$, $Y_{lf2}(s) = 1$

First conditional score function for $h(s)$ is

$$U_{0,h(s)} = \begin{cases} \dfrac{\mathbb{1}(T_D \geq s)}{\sum\limits_{k:X_2<t_k\leq T_D} h(t_k)\mu(t_k)}\mu(s) - \mu(s)U\mathbb{1}(T_D \geq s), & T_D \leq t_{lf2} \\ \dfrac{1}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}\mu(s) - \mu(s)U, & T_D = \infty. \end{cases}$$

Next the numerator $E[U_{0,h(s)}L_0]$ can be calculated as

$$\begin{aligned} &E[U_{0,h(s)}L_0] \\ &= \mu(s)E[\frac{\mathbb{1}(s \leq T_D \leq t_{lf2})}{\sum\limits_{k:X_2<t_k\leq T_D} h(t_k)\mu(t_k)}L_0] - \mu(s)E[\mathbb{1}(s \leq T_D \leq t_{lf2})UL_0] \\ &\quad + \frac{\mu(s)}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}E[\mathbb{1}(T_D = \infty)L_0] - \mu(s)E[\mathbb{1}(T_D = \infty)UL_0]. \end{aligned}$$

The term $E[\frac{\mathbb{1}(s \leq T_D \leq t_{lf2})}{\sum\limits_{k:X_2<t_k\leq T_D} h(t_k)\mu(t_k)}L_0]$ can be derived as

$$\begin{aligned} &\int_0^\infty \int_{s^-}^{t_{lf2}} \frac{\sum\limits_{k:X_2<t_k\leq t_D} h(t_k)\mu(t_k)}{\sum\limits_{k:X_2<t_k\leq t_D} h(t_k)\mu(t_k)} \exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}\eta dH(t_D)du \\ &= \int_{s^-}^{t_{lf2}} \frac{1}{1 + \sum\limits_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)}\eta e^{-\eta H(t_D)}dH(t_D) = U(s^-, t_{lf2}). \end{aligned}$$

The term $E[\mathbb{1}(s \leq T_D \leq t_{lf2})UL_0]$ is derived as

$$\begin{aligned} &E[\mathbb{1}(s \leq T_D \leq t_{lf2})UL_0] \\ &= \int_0^\infty \int_{s^-}^{t_{lf2}} [\sum_{k:X_2<t_k\leq t_D} h(t_k)\mu(t_k)]u \exp\{-\eta H(t_D) - u[1 + \sum_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]\}\eta dH(t_D)du \\ &= \int_{s^-}^{t_{lf2}} \frac{\sum\limits_{k:X_2<t_k\leq t_D} h(t_k)\mu(t_k)}{[1 + \sum\limits_{k:0<t_k\leq t_D} h(t_k)\mu(t_k)]^2}\eta e^{-\eta H(t_D)}dH(t_D) = U(s^-, t_{lf2}) - S_2. \end{aligned}$$

The term $\frac{1}{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)} E[\mathbb{1}(T_D=\infty)L_0]$ is derived as

$$\frac{1}{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)} E[\mathbb{1}(T_D=\infty)L_0]$$
$$= \frac{1}{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)} \int_0^\infty [\sum_{k:t_k>X_2} h(t_k)\mu(t_k)] \exp\{-\eta H(t_{lf2}) - u[1+\sum_{k:t_k>0} h(t_k)\mu(t_k)]\} du$$
$$= \frac{1}{1+\sum_{k:t_k>0} h(t_k)\mu(t_k)} e^{-\eta H(t_{lf2})} = V.$$

The term $E[\mathbb{1}(T_D=\infty)UL_0]$ is derived as

$$E[\mathbb{1}(T_D=\infty)UL_0]$$
$$= \int_0^\infty [\sum_{k:t_k>X_2} h(t_k)\mu(t_k)] u \exp\{-\eta H(t_{lf2}) - u[1+\sum_{k:t_k>0} h(t_k)\mu(t_k)]\} du$$
$$= \frac{\sum_{k:t_k>X_2} h(t_k)\mu(t_k)}{[1+\sum_{k:t_k>0} h(t_k)\mu(t_k)]^2} e^{-\eta H(t_{lf2})} = V - T.$$

The numerator $E[U_{0,h(s)}L_0]$ is calculated as

$$E[U_{0,h(s)}L_0] = \mu(s)[U(s^-, t_{lf2}) - [U(s^-, t_{lf2}) - S_2] + V - (V-T)]$$
$$= \mu(s)(S_2 + T).$$

Thus, unconditional score function $U_{h(s)}$ is derived as

$$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = \mu(s)\frac{S_2+T}{W(X_2, t_{lf2}) + Z}.$$

- $s > t_{lf2}$, i.e. $Y_{lf2}(s) = 0$

First, conditional score function $U_{0,h(s)}$ is

$$U_{0,h(s)} = \begin{cases} 0, & T_D \leq t_{lf2} \\ \dfrac{1}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}\mu(s) - \mu(s)U, & T_D = \infty. \end{cases}$$

Next the numerator $E[U_{0,h(s)}L_0]$ can be calculated as

$$\begin{aligned} & E[U_{0,h(s)}L_0] \\ = & \frac{\mu(s)}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)} E[\mathbb{1}(T_D = \infty)L_0] - \mu(s)E[\mathbb{1}(T_D = \infty)UL_0] \\ = & \mu(s)(V + V - T) = \mu(s)[V - (V - T)] = \mu(s)T. \end{aligned}$$

Thus, unconditional score function $U_{h(s)}$ is derived as

$$U_{h(s)} = \frac{E[U_{0,h(s)}L_0]}{E[L_0]} = \mu(s)\frac{T}{W(X_2, t_{lf2}) + Z}.$$

Combining these results, the contribution of a subject, with both intermediate and terminal events censored at X_2, to the unconditional score function, can be written as

$$\begin{aligned} U_h(s) = & \mu(s)\Bigg\{[\mathbb{1}(X_2 \geq t_{lf1}) + \mathbb{1}(t_{lf2} \leq X_2 < t_{lf1})Y_2(s)]\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} \\ & + \mathbb{1}(t_{lf2} \leq X_2 < t_{lf1})[1 - Y_2(s)]\Big[\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} - \frac{1}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}\Big] \\ & + \mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\frac{-T + Y_2(s)(R + V) - [1 - Y_2(s)]Y_{lf2}(s)S_2}{W(X_2, t_{lf2}) + Z}\Bigg\}. \end{aligned} \tag{B.6}$$

Combining equations (B.4), (B.5) and (B.6), we have unconditional score function for $h(s)$ as

$$U_{h^{(k+1)}}(s) = \frac{dN_1(s)}{h^{(k+1)}(s)} - \Psi_h^{(k)}(s),$$

where

$$\begin{aligned}
\Psi_h(s) =& \Delta_2 Y_2(s)\mu(s)\frac{1}{\sum\limits_{k:0<t_k\le X_2} h(t_k)\mu(t_k)} \\
&+(1-\Delta_1)\mu(s)\Bigg\{[\mathbb{1}(X_2 \ge t_{lf1}) + \mathbb{1}(t_{lf2} \le X_2 < t_{lf1})Y_2(s)]\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} \\
&\quad + \mathbb{1}(t_{lf2} \le X_2 < t_{lf1})[1-Y_2(s)]\Big[\frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} - \frac{1}{\sum\limits_{k:t_k>X_2} h(t_k)\mu(t_k)}\Big] \\
&\quad + \mathbb{1}(X_2 < t_{lf1})\mathbb{1}(X_2 < t_{lf2})\frac{-T + Y_2(s)(R+V) - [1-Y_2(s)]Y_{lf2}(s)S_2}{W(X_2, t_{lf2}) + Z}\Bigg\} \\
&+\Delta_1(1-\Delta_2)\mu(s)\mathbb{1}(X_2 < t_{lf2})\Bigg\{ - \frac{1}{\sum\limits_{k:t_k>0} h(t_k)\mu(t_k)} \\
&\quad + \frac{Q + Y_2(s)P(X_2, t_{lf2}) + [1-Y_2(s)]Y_{lf2}(s)P(s^-, t_{lf2})}{U(X_2, t_{lf2}) + V}\Bigg\}.
\end{aligned}$$

B.3.2.3 M step

Suppose we have n independent subjects with observed data $(X_{1i}, \Delta_{1i}, X_{2i}, \Delta_{2i})$, $i = 1, 2, ..., n$. The estimator of $h^{(k+1)}(s)$ can be derived by solving $\sum\limits_{i=1}^{n} U_{i,h}(s) = 0$. Now we want to solve

$$\sum_{i=1}^{n}\left\{\frac{dN_{1i}(s)}{h^{(k+1)}(s)} - \Psi_{h,i}(s)\right\} = 0,$$

which gives

$$h^{(k+1)}(s) = \frac{\sum\limits_i dN_{1i}(s)}{\sum\limits_i \Psi_{h,i}^{(k)}(s)}.$$

The above equation solves for $h(s)$ iteratively until convergence, and the estimator at convergence, $\hat{h}(s)$, is a consistent estimator of $h(s)$ (Tsodikov (2003)).

APPENDIX C

A Mechanistic Joint Model to Investigate the Screening Effect on Cancer Mortality

C.1 Complete-data Likelihood

The complete-data likelihood L_{com} for a single subject is:

$$L_{com} = L_0 f_1(t_1),$$

where L_0 is the conditional likelihood of observed data, given T_1, which can be seen from (4.4), (4.5) and (4.6); and $f_1(t_1)$ is the distribution of T_1, $f_1(t_1) = \mu_1 dH_I(t_1) e^{-\mu_1 H_I(t_1)}$. Then we can obtain L_{com} in the following three cases:

1. Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$L_{com} = \begin{cases} \eta\mu_1 dH_I(X_I) dH_D(X_D) e^{-(\mu_1+\mu_2)H_I(X_I) - \eta[H_D(X_D) - H_D(X_I)]}, & T_1 = X_I \\ \eta\mu_1\mu_2 dH_I(t_1) dH_I(X_I) dH_D(X_D) e^{-\mu_1 H_I(t_1) - \mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(t_1)]}, & T_1 \in (X_I, X_D] \end{cases}$$

2. Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$L_{com} = \begin{cases} \mu_1 dH_I(X_I) e^{-(\mu_1+\mu_2)H_I(X_I)-\eta[H_D(X_D)-H_D(X_I)]}, & T_1 = X_I \\ \mu_1\mu_2 dH_I(t_1) dH_I(X_I) e^{-\mu_1 H_I(t_1)-\mu_2 H_I(X_I)-\eta[H_D(X_D)-H_D(t_1)]}, & T_1 \in (X_I, X_D] \\ \mu_1\mu_2 dH_I(X_I) dH_I(t_1) e^{-\mu_1 H_I(t_1)-\mu_2 H_I(X_I)}, & T_1 > X_D. \end{cases}$$

3. Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$L_{com} = \mu_1 dH_I(t_1) e^{-\mu_1 H_I(t_1)-\mu_2 H_I(X_I)}, \quad T_1 > X_I.$$

C.2 Prediction of Survival Function for Causal Incidence

Prediction of time to causal incidence is of interest in analysis. In this section, we derive the survival functions for causal incidence, given observed data and estimates of $\eta, \mu_1, \mu_2, \{dH_I\}$ and $\{dH_D\}$:

$$\begin{aligned} & S_{T_1}(t_1|Observed\ \ Data) \\ &= P(T_1 > t_1|Observed\ \ Data) \\ &= \frac{P(T_1 > t_1, Observed\ \ Data)}{P(Observed\ \ Data)} \\ &= \frac{\int_{t_1}^{\infty} L_{com} dT_1}{L_{obs}}. \end{aligned}$$

Note the denominator is the observed data likelihood (4.10), and L_{com} in the numerator is the complete data likelihood in Appendix C.1. The survival function can be derived in the following three conditions:

1. Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$S_{T_1}(t_1|X_I, X_D, \Delta_I = 1, \Delta_D = 1) = \frac{\int_{t_1}^{\infty} L_{com} dT_1}{L_{11}},$$

where L_{11} is the observed likelihood for $\Delta_I = 1$, $\Delta_D = 1$ as (4.7).

- If $t_1 < X_I$,

$$\int_{t_1}^{\infty} L_{com} dT_1 = \mathbb{1}(T_1 = X_I) + \int_{X_I}^{X_D} L_{com} dT_1 = L_{11}.$$

- If $t_1 \in [X_I, X_D]$,

$$\begin{aligned}
&\int_{t_1}^{\infty} L_{com} dT_1 = \int_{t_1}^{X_D} L_{com} dT_1 \\
&= \int_{t_1}^{X_D} \eta\mu_1\mu_2 dH_I(X_I) dH_D(X_D) e^{-\mu_1 H_I(x) - \mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(x)]} dH_I(x) \\
&= \eta\mu_1\mu_2 dH_I(X_I) dH_D(X_D) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} \int_{t_1}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x).
\end{aligned}$$

- If $t_1 > X_D$,

$$\int_{t_1}^{\infty} L_{com} dT_1 = 0.$$

After some algebra, we obtain the conditional survival function for causal incidence T_1, given observed incidence at X_I and observed death at X_D as:

$$\begin{aligned}
&S_{T_1}(t_1 | X_I, X_D, \Delta_I = 1, \Delta_D = 1) \\
&= \begin{cases}
1, & t_1 < X_I \\
\frac{\mu_2 \int_{t_1}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x)}{e^{\eta H_D(X_I) - \mu_1 H_I(X_I)} + \mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x)}, & t_1 \in [X_I, X_D] \\
0, & t_1 > X_D.
\end{cases}
\end{aligned}$$

2. Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$S_{T_1}(t_1|X_I, X_D, \Delta_I = 1, \Delta_D = 0) = \frac{\int_{t_1}^{\infty} L_{com} dT_1}{L_{10}},$$

where L_{10} is the observed likelihood for $\Delta_I = 1$, $\Delta_D = 0$ as (4.8).

- If $t_1 < X_I$,

$$\int_{t_1}^{\infty} L_{com} dT_1 = \mathbb{1}(T_1 = X_I) + \int_{X_I}^{\infty} L_{com} dT_1 = L_{10}.$$

- If $t_1 \in [X_I, X_D]$,

$$\begin{aligned}
\int_{t_1}^{\infty} L_{com} dT_1 &= \int_{t_1}^{X_D} L_{com} dT_1 + \int_{X_D}^{\infty} L_{com} dT_1 \\
&= \int_{t_1}^{X_D} \mu_1 \mu_2 dH_I(X_I) e^{-\mu_1 H_I(t_1) - \mu_2 H_I(X_I) - \eta[H_D(X_D) - H_D(t_1)]} dH_I(t_1) \\
&\quad + \int_{X_D}^{\infty} \mu_1 \mu_2 dH_I(X_I) e^{-\mu_1 H_I(t_1) - \mu_2 H_I(X_I)} dH_I(t_1) \\
&= dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} \Bigg[\mu_1 \mu_2 \int_{t_1}^{X_D} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) \\
&\qquad\qquad + \mu_2 e^{\eta H_D(X_D) - \mu_1 H_I(X_D)} \Bigg].
\end{aligned}$$

- If $t_1 > X_D$,

$$\begin{aligned}
\int_{t_1}^{\infty} L_{com} dT_1 &= \int_{t_1}^{\infty} \mu_1 \mu_2 dH_I(X_I) e^{-\mu_1 H_I(t_1) - \mu_2 H_I(X_I)} dH_I(t_1) \\
&= \mu_2 dH_I(X_I) e^{-\mu_1 H_I(t_1) - \mu_2 H_I(X_I)}.
\end{aligned}$$

Therefore, we can obtain the conditional survival function for causal incidence

T_1, given observed incidence at X_I and terminal event censored at X_D as:

$$S_{T_1}(t_1|X_I, X_D, \Delta_I = 1, \Delta_D = 0)$$

$$= \begin{cases} 1, t_1 < X_I \\ \frac{\mu_1\mu_2 \int_{t_1}^{X_D} e^{\eta H_D(x)-\mu_1 H_I(x)} dH_I(x)+\mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(X_D)}}{\mu_1 e^{\eta H_D(X_I)-\mu_1 H_I(X_I)}+\mu_1\mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x)-\mu_1 H_I(x)} dH_I(x)+\mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(X_D)}}, & t_1 \in [X_I, X_D] \\ \frac{\mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(t_1)}}{\mu_1 e^{\eta H_D(X_I)-\mu_1 H_I(X_I)}+\mu_1\mu_2 \int_{X_I}^{X_D} e^{\eta H_D(x)-\mu_1 H_I(x)} dH_I(x)+\mu_2 e^{\eta H_D(X_D)-\mu_1 H_I(X_D)}}, & t_1 > X_D. \end{cases}$$

3. Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$S_{T_1}(t_1|X_I, X_D, \Delta_I = 0, \Delta_D = 0) = \frac{\int_{t_1}^{\infty} L_{com} dT_1}{L_{00}},$$

where L_{00} is the observed likelihood for $\Delta_I = 0$, $\Delta_D = 0$ as (4.9).

- If $t_1 < X_I$,
$$\int_{t_1}^{\infty} L_{com} dT_1 = \int_{X_I}^{\infty} L_{com} dT_1 = L_{00}.$$

- If $t_1 \geq X_I$,

$$\int_{t_1}^{\infty} L_{com} dT_1 = \int_{t_1}^{\infty} \mu_1 e^{-\mu_1 H_I(t_1)-\mu_2 H_I(X_I)} dH_I(t_1) = e^{-\mu_1 H_I(t_1)-\mu_2 H_I(X_I)}.$$

Then, the conditional survival function for causal incidence T_1, given neither event observed before X_I is:

$$S_{T_1}(t_1|X_I, X_D, \Delta_I = 0, \Delta_D = 0) = \begin{cases} 1, & t_1 < X_I \\ e^{-\mu_1[H_I(t_1)-H_I(X_I)]}, & t_1 \geq X_I. \end{cases}$$

C.3 Derivation of Hazard Terms

For ease of notation, we suppress the subscript i for each individual subject.

1. The hazard of having incidence at time t:

$$d\Lambda_I(t) = \Theta_I(t)dH_I(t) = Pr\big(dN_I(t) = 1|Y_I(t) = 1\big) = (\mu_1 + \mu_2)dH_I(t)$$
$$\Rightarrow \Theta_I(t) = \mu_1 + \mu_2.$$

2. The hazard of dying from cancer at time t, given incidence at time t_I:

$$\begin{aligned} d\Lambda_D(t;t_I) = \Theta_D(t;t_I)dH_D(t) &= Pr\big(dN_D(t) = 1|Y_D(t) = 1, dN_I(t_I) = 1\big) \\ &= \frac{Pr(T_D = t, T_I = t_I)}{Pr(T_D \geq t, T_I = t_I)} = \frac{L_{11}(t_I,t)}{L_{10}(t_I,t)} \\ \Rightarrow \Theta_D(t;t_I) &= \frac{L_{11}(t_I,t)}{L_{10}(t_I,t)dH_D(t)} \\ &= \eta\mu_1 \frac{e^{\eta H_D(t_I) - \mu_1 H_I(t_I)} + \mu_2 \int_{t_I}^{t} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x)}{\mu_1 e^{\eta H_D(t_I) - \mu_1 H_I(t_I)} + \mu_1\mu_2 \int_{t_I}^{t} e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x) + \mu_2 e^{\eta H_D(t) - \mu_1 H_I(t)}}. \end{aligned}$$

C.4 Derivation of Score Function

Based on the log-likelihood expression (4.11), we have

$$\begin{aligned} l_I &= \sum_{i=1}^{n} l_{Ii} = \sum_{i=1}^{n} \int_0^\tau \Big[\log \Theta_{Ii}(t) + \log dH_I(t)\Big] dN_{Ii}(t) - Y_{Ii}(t)\Theta_{Ii}(t)dH_I(t) \\ &= \sum_{i=1}^{n} \int_0^\tau \Big\{dN_{Ii}(t) \log d\Lambda_{Ii}(t) - Y_{Ii}(t)d\Lambda_{Ii}(t)\Big\}, \\ l_D &= \sum_{i=1}^{n} l_{Di} = \sum_{i=1}^{n} \int_0^\tau \Big[\log \Theta_{Di}(t) + \log dH_D(t)\Big] dN_{Di}(t) - Y_{Di}(t)\Theta_{Di}(t)dH_D(t) \\ &= \sum_{i=1}^{n} \int_0^\tau \Big\{dN_{Di}(t) \log d\Lambda_{Di}(t) - Y_{Di}(t)d\Lambda_{Di}(t)\Big\}, \end{aligned}$$

where $d\Lambda_{Ii}(t) = \Theta_{Ii}(t)dH_I(t)$ and $d\Lambda_{Di}(t) = \Theta_{Di}(t)dH_D(t)$.

Taking derivatives of l, with respect to β, $dH_I(s)$, and $dH_D(s)$, respectively, we can obtain the corresponding score functions. Since $l = l_I + l_D$, score function $U = U_I + U_D$.

Next we derive $U_{I,(\cdot)}$. Note $(\cdot)$ can be β, $dH_I(s)$, and $dH_D(s)$:

$$\begin{aligned}
U_{I,(\cdot)} = \frac{\partial l_I}{\partial(\cdot)} &= \frac{\partial}{\partial(\cdot)} \sum_{i=1}^{n} \int_0^{\infty} \Big\{ dN_{Ii}(t) \log d\Lambda_{Ii}(t) - Y_{Ii}(t) d\Lambda_{Ii}(t) \Big\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \Big\{ \frac{\partial \log d\Lambda_{Ii}(t)}{\partial(\cdot)} dN_{Ii}(t) - Y_{Ii}(t) d\Lambda_{Ii}(t) \frac{\partial \log d\Lambda_{Ii}(t)}{\partial(\cdot)} \Big\} \\
&= \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \log d\Lambda_{Ii}(t)}{\partial(\cdot)} dM_{Ii}(t) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \frac{1}{d[\Theta_{Ii}(t)dH_I(t)]} \frac{\partial [\Theta_{Ii}(t)dH_I(t)]}{\partial(\cdot)} dM_{Ii}(t) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \frac{1}{d[\Theta_{Ii}(t)dH_I(t)]} \Big[dH_I(t) \frac{\partial \Theta_{Ii}(t)}{\partial(\cdot)} + \Theta_{Ii}(t) \frac{\partial dH_I(x)}{\partial(\cdot)} \Big] dM_{Ii}(t) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \Big[\frac{1}{\Theta_{Ii}(t)} \frac{\partial \Theta_{Ii}(t)}{\partial(\cdot)} + \frac{1}{dH_I(t)} \frac{\partial dH_I(t)}{\partial(\cdot)} \Big] dM_{Ii}(t).
\end{aligned}$$

Since $\dfrac{\partial \Theta_I(t)}{\partial dH_I(s)} = \mathbb{1}(t \geq s) \dfrac{\partial \Theta_I(t)}{\partial dH_I(s)}$ and $\dfrac{\partial dH_I(t)}{\partial dH_I(s)} = \mathbb{1}(t = s)$, then

$$U_{I,dH_I(s)} = \sum_{i=1}^{n} \Big\{ \frac{dM_{Ii}(s)}{dH_I(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Ii,dH_I(s)}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t) \Big\}.$$

Since $\dfrac{\partial \Theta_I(t)}{\partial dH_D(s)} = \mathbb{1}(t \geq s) \dfrac{\partial \Theta_I(t)}{\partial dH_D(s)}$ and $\dfrac{\partial dH_I(t)}{\partial dH_D(s)} = 0$, then

$$U_{I,dH_D(s)} = \sum_{i=1}^{n} \int_s^{\tau} \frac{\dot{\Theta}_{Ii,dH_D(s)}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t).$$

Since $\dfrac{\partial dH_I(t)}{\partial \beta} = 0$, then

$$U_{I,\beta} = \sum_{i=1}^{n} \int_0^{\tau} \frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t).$$

Similarly, we can have $U_{D,(\cdot)}$ as

$$\begin{aligned}
U_{D,dH_I(s)} &= \sum_{i=1}^{n} \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_I(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t), \\
U_{D,dH_D(s)} &= \sum_{i=1}^{n} \left\{ \frac{dM_{Di}(s)}{dH_D(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_D(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \\
U_{D,\beta} &= \sum_{i=1}^{n} \int_0^{\tau} \frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)} dM_{Di}(t).
\end{aligned}$$

Finally, since $U_{(\cdot)} = U_{I,(\cdot)} + U_{D,(\cdot)}$, and in our model, $\dot{\Theta}_{Ii,dH_I(s)}(t) = \dot{\Theta}_{Ii,dH_D(s)}(t) = 0$, score functions for $dH_I(s)$, $dH_D(s)$, and β, respectively, can be written as

$$\begin{aligned}
U_{dH_I(s)} &= \sum_{i=1}^{n} \left\{ \frac{dM_{Ii}(s)}{dH_I(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_I(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \\
U_{dH_D(s)} &= \sum_{i=1}^{n} \left\{ \frac{dM_{Di}(s)}{dH_D(s)} + \int_s^{\tau} \frac{\dot{\Theta}_{Di,dH_D(s)}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}, \\
U_{\beta} &= \sum_{i=1}^{n} \int_0^{\tau} \left\{ \frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t) + \frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}.
\end{aligned}$$

C.5 EM Algorithm

This section presents the derivation of EM algorithm to estimate the baseline hazards of both intermediate and terminal events for our model, using the methods of Tsodikov (2003):

Conditional score function can be written as

$$U_0(s) = \frac{\partial log L_{com}}{\partial dH(s)}.$$

In the (k+1)th iteration,

E-step: Unconditional score function

$$U(s) = E[U_0(s)||L_0^{(k)}] = \frac{E[U_0(s)L_0^{(k)}]}{E[L_0^{(k)}]}.$$

M-step: Set $U(s) = 0$, then we can obtain $dH^{(k+1)}(s)$ to update the expression of $dH(s)$.

The rest of this section is organized as follows. First, we write out the conditional score function for baseline hazard of incidence $\{dH_I\}$, for the three different cases respectively (both events observed, incidence observed yet death censored, and neither events observed). Next we derive the E step for $\{dH_I\}$. Finally we derive the M step to maximize the likelihood with respect to $\{dH_I\}$, which has a closed-form expression analogous to the weighted Breslow-type estimators of Chen (2009). Similar steps are taken to derive the EM algorithm for baseline hazard of death $\{dH_D\}$. The following notations are introduced

$$V(a,b) = \int_a^b e^{\eta H_D(x) - \mu_1 H_I(x)} dH_I(x),$$

$$W(s) = e^{\eta H_D(s) - \mu_1 H_I(s)}.$$

C.5.1 Conditional score functions

1. Conditional score function w.r.t $\{dH_I\}$ is derived as

$$U_{0,dH_I}(s) = \frac{\partial log L_{com}}{\partial dH_I(s)}.$$

- Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$U_{0,dH_I}(s) = \begin{cases} \frac{\mathbb{1}(X_I=s)}{dH_I(s)} - (\mu_1+\mu_2)\mathbb{1}(X_I \geq s), & T_1 = X_I \\ \frac{\mathbb{1}(T_1=s)}{dH_I(s)} + \frac{\mathbb{1}(X_I=s)}{dH_I(s)} - \mu_1\mathbb{1}(T_1 \geq s) - \mu_2\mathbb{1}(X_I \geq s), & T_1 \in (X_I, X_D]. \end{cases}$$

- Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$U_{0,dH_I}(s) = \begin{cases} \frac{\mathbb{1}(X_I=s)}{dH_I(s)} - (\mu_1+\mu_2)\mathbb{1}(X_I \geq s), & T_1 = X_I \\ \frac{\mathbb{1}(T_1=s)}{dH_I(s)} + \frac{\mathbb{1}(X_I=s)}{dH_I(s)} - \mu_1\mathbb{1}(T_1 \geq s) - \mu_2\mathbb{1}(X_I \geq s), & T_1 > X_I. \end{cases}$$

- Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$U_{0,dH_I}(s) = \frac{\mathbb{1}(T_1 = s)}{dH_I(s)} - \mu_1\mathbb{1}(T_1 \geq s) - \mu_2\mathbb{1}(X_I \geq s), \quad T_1 > X_I.$$

2. Similarly, conditional score function w.r.t $\{dH_D\}$:

$$U_{0,dH_D}(s) = \frac{\partial log L_{com}}{\partial dH_D(s)}.$$

- Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$):

$$U_{0,dH_D}(s) = \begin{cases} \frac{\mathbb{1}(X_D=s)}{dH_D(s)} - \eta[\mathbb{1}(X_D \geq s) - \mathbb{1}(X_I \geq s)], & T_1 = X_I \\ \frac{\mathbb{1}(X_D=s)}{dH_D(s)} - \eta[\mathbb{1}(X_D \geq s) - \mathbb{1}(T_1 \geq s)], & T_1 \in (X_I, X_D]. \end{cases}$$

- Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$):

$$U_{0,dH_D}(s) = \begin{cases} -\eta[\mathbb{1}(X_D \geq s) - \mathbb{1}(X_I \geq s)], & T_1 = X_I \\ -\eta[\mathbb{1}(X_D \geq s) - \mathbb{1}(T_1 \geq s)], & T_1 \in (X_I, X_D] \\ 0, & T_1 > X_D. \end{cases}$$

- Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$):

$$U_{0,dH_D}(s) = 0.$$

C.5.2 E step

The unconditional score functions are

$$U_{dH_I}(s) = E[U_{0,dH_I}(s)||L_0^{(k)}] = \frac{E[U_{0,dH_I}(s)L_0^{(k)}]}{E[L_0^{(k)}]}$$
$$U_{dH_D}(s) = E[U_{0,dH_D}(s)||L_0^{(k)}] = \frac{E[U_{0,dH_D}(s)L_0^{(k)}]}{E[L_0^{(k)}]}$$

In the rest of EM algorithm section, we keep iteration index for $(k+1)^{th}$ iteration when needed, and drop index for k^{th} for brevity. Note that the denominator $E[L_0^{(k)}]$ is just observed data likelihood, as (4.10).

1. Subject has incidence at X_I, and dies at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 1$)

   First, the denominator $E[L_0^{(k)}]$ is the observed data likelihood (4.7), and can be rewritten as

$$E[L_0] = L_{obs,11} = \eta\mu_1 dH_I(X_I)dH_D(X_D)e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)}[W(X_I) + \mu_2 V(X_I, X_D)].$$

   Next, we calculate the numerators $E[U_{0,dH_I}(s)L_0]$ and $E[U_{0,dH_D}(s)L_0]$, and finally obtain the unconditional score functions $U_{dH_I}(s)$ and $U_{dH_D}(s)$. Since $U_{0,dH_I}(s)$

and $U_{0,dH_D}(s)$ depend on s, we consider the calculation in three cases, $s \leq X_I$, $X_I < s \leq X_D$, and $s > X_D$, respectively.

(a) $s \leq X_I$, i.e. $Y_I(s) = 1$, $Y_D(s) = 1$

Based on conditional score functions derived in Appendix C.5.1,

- To calculate $U_{dH_I}(s)$, first

$$U_{0,dH_I}(s) = \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - (\mu_1 + \mu_2),$$

since $U_{0,dH_I}(s)$ does not depend on the unobserved T_1,

$$E[U_{0,dH_I}(s)L_0] = U_{0,dH_I}(s)E[L_0] = \Big[\frac{dN_I(s)}{dH_I^{(k+1)}(s)} - (\mu_1 + \mu_2)\Big]E[L_0].$$

Thus, the unconditional score function for $dH_I(s)$,

$$U_{dH_I}(s) = \frac{E[U_{0,dH_I}(s)L_0]}{E[L_0]} = \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - (\mu_1 + \mu_2).$$

- Similarly, we can obtain that $U_{0,dH_D}(s) = 0$, then $E[U_{0,dH_D}(s)L_0] = 0$. Thus, the unconditional score function for $dH_D(s)$,

$$U_{dH_D}(s) = 0.$$

(b) $X_I < s \leq X_D$, i.e. $Y_I(s) = 0$, $Y_D(s) = 1$

- To calculate $U_{dH_I}(s)$, first

$$U_{0,dH_I}(s) = \begin{cases} 0, & T_1 = X_I \\ \frac{\mathbb{1}(T_1=s)}{dH_I^{(k+1)}(s)} - \mu_1 \mathbb{1}(T_1 \geq s), & T_1 \in (X_I, X_D]. \end{cases}$$

Then the numerator

$$E[U_{0,dH_I}(s)L_0] = \frac{E[\mathbb{1}(T_1 = s)L_0]}{dH_I^{(k+1)}(s)} - \mu_1 E[\mathbb{1}(T_1 \geq s)L_0], \;\; T_1 \in (X_I, X_D].$$

Expression of conditional likelihood L_0 is given as (4.4) in Section 4.2.3, then we can derive the term $E[\mathbb{1}(T_1 = s)L_0]$ as

$$\begin{aligned}
&E[\mathbb{1}(T_1 = s)L_0] \\
=&\eta\mu_1\mu_2 dH_I(s)dH_I(X_I)dH_D(X_D)e^{-\eta H_D(X_D)-\mu_2 H_I(X_I)+\eta H_D(s)-\mu_1 H_I(s)} \\
=&\eta\mu_1\mu_2 dH_I(s)dH_I(X_I)dH_D(X_D)e^{-\eta H_D(X_D)-\mu_2 H_I(X_I)}W(s).
\end{aligned}$$

The term $E[\mathbb{1}(T_1 \geq s)L_0]$ can be derived as

$$\begin{aligned}
&E[\mathbb{1}(T_1 \geq s)L_0] \\
=&\int_{s^-}^{X_D} \eta\mu_1\mu_2 dH_I(X_I)dH_D(X_D)e^{-\eta H_D(X_D)-\mu_2 H_I(X_I)+\eta H_D(x)-\mu_1 H_I(x)}dH_I(x) \\
=&\eta\mu_1\mu_2 dH_I(X_I)dH_D(X_D)e^{-\eta H_D(X_D)-\mu_2 H_I(X_I)}V(s^-, X_D).
\end{aligned}$$

Then,

$$\begin{aligned}
E[U_{0,dH_I}(s)L_0] = \eta\mu_1\mu_2 dH_I(X_I)dH_D(X_D)e^{-\eta H_D(X_D)-\mu_2 H_I(X_I)} \\
\left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)}W(s) - \mu_1 V(s^-, X_D)\right].
\end{aligned}$$

Thus, the unconditional score function for $dH_I(s)$

$$U_{dH_I(s)} = \mu_2 \frac{\dfrac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)}W(s) - \mu_1 V(s^-, X_D)}{W(X_I) + \mu_2 V(X_I, X_D)}.$$

- To calculate $U_{dH_D}(s)$, first

$$U_{0,dH_D}(s) = \begin{cases} \dfrac{\mathbb{1}(X_D = s)}{dH_D^{(k+1)}(s)} - \eta, & T_1 = X_I \\ \dfrac{\mathbb{1}(X_D = s)}{dH_D^{(k+1)}(s)} - \eta + \eta\mathbb{1}(T_1 \geq s), & T_1 \in (X_I, X_D]. \end{cases}$$

Then the numerator

$$E[U_{0,dH_D}(s)L_0] = \left[\frac{\mathbb{1}(X_D = s)}{dH_D^{(k+1)}(s)} - \eta\right] E[L_0] + \eta E[\mathbb{1}(T_1 \geq s)L_0]$$

$$= \left[\frac{\mathbb{1}(X_D = s)}{dH_D^{(k+1)}(s)} - \eta\right] E[L_0]$$

$$+\eta^2 \mu_1 \mu_2 dH_I(X_I) dH_D(X_D) e^{-\eta H_D(X_D) - \mu_2 H_I(X_I)} V(s^-, X_D).$$

Thus, the unconditional score function for $dH_D(s)$

$$U_{dH_D(s)} = \frac{dN_D(s)}{dH_D^{(k+1)}(s)} - \eta \frac{W(X_I) + \mu_2 V(X_I, s^-)}{W(X_I) + \mu_2 V(X_I, X_D)}.$$

(c) $s > X_D$, i.e. $Y_I(s) = 0$, $Y_D(s) = 0$

$$U_{0,dH_I}(s) = U_{0,dH_D}(s) = 0.$$

Thus, the unconditional score functions

$$U_{dH_I}(s) = U_{dH_D}(s) = 0.$$

Combining these results, the contribution of a subject, with both intermediate and terminal events observed at X_I and X_D, to the unconditional score functions

can be written as

$$
\begin{aligned}
U_{dH_I}(s) = & \frac{dN_I(s)}{dH_I^{(k+1)}(s)} \\
& + \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right] Y_D(s)[1 - Y_I(s)] \frac{\mu_2 W(s)}{W(X_I) + \mu_2 V(X_I, X_D)} \\
& - \left\{ Y_I(s)(\mu_1 + \mu_2) + Y_D(s)[1 - Y_I(s)]\mu_2 \frac{\mu_1 V(s^-, X_D) - W(s)}{W(X_I) + \mu_2 V(X_I, X_D)} \right\}, \\
U_{dH_{D(s)}} = & \frac{dN_D(s)}{dH_D^{(k+1)}(s)} - Y_D(s)[1 - Y_I(s)]\eta \frac{W(X_I) + \mu_2 V(X_I, s^-)}{W(X_I) + \mu_2 V(X_I, X_D)}. \qquad \text{(C.1)}
\end{aligned}
$$

2. Subject has incidence at X_I, and is censored at X_D (i.e. $\Delta_I = 1$, $\Delta_D = 0$)

   We now proceed through the same steps as for the case $\Delta_I = \Delta_D = 1$. In this case, the denominator $E[L_0^{(k)}]$ is the observed data likelihood (4.8), and can be rewritten as

$$
\begin{aligned}
E[L_0] &= L_{obs,10} \\
&= dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} [\mu_1 W(X_I) + \mu_1 \mu_2 V(X_I, X_D) + \mu_2 W(X_D)].
\end{aligned}
$$

   Then, we calculate $U_{dH_I}(s)$ and $U_{dH_D}(s)$ in three cases, $s \leq X_I$, $X_I < s \leq X_D$, and $s > X_D$, respectively.

   (a) $s \leq X_I$, i.e. $Y_I(s) = 1$, $Y_D(s) = 1$

   - To calculate $U_{dH_I}(s)$, first

$$
U_{0,dH_I}(s) = \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - (\mu_1 + \mu_2),
$$

     which does not depend on the unobserved T_1.

Thus, the unconditional score function for $dH_I(s)$,

$$U_{dH_I}(s) = \frac{E[U_{0,dH_I}(s)L_0]}{E[L_0]} = U_{0,dH_I}(s)\frac{E[L_0]}{E[L_0]} = \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - (\mu_1 + \mu_2).$$

- To calculate $U_{dH_D}(s)$, since $U_{0,dH_D}(s) = 0$, then $E[U_{0,dH_D}(s)L_0] = 0$. Thus, the unconditional score function for $dH_D(s)$,

$$U_{dH_D}(s) = 0.$$

(b) $X_I < s \leq X_D$, i.e. $Y_I(s) = 0$, $Y_D(s) = 1$

- To calculate $U_{dH_I}(s)$, first

$$U_{0,dH_I}(s) = \begin{cases} 0, & T_1 = X_I, \\ \frac{\mathbb{1}(T_1=s)}{dH_I^{(k+1)}(s)} - \mu_1 \mathbb{1}(T_1 \geq s), & T_1 \in (X_I, X_D], \\ -\mu_1, & T_1 > X_D. \end{cases}$$

Then the numerator

$$E[U_{0,dH_I}(s)L_0] = \frac{E[\mathbb{1}(T_1 = s)\mathbb{1}(X_I < T_1 \leq X_D)L_0]}{dH_I^{(k+1)}(s)} - \mu_1 E[\mathbb{1}(s \leq T_1 \leq X_D)L_0] - \mu_1 E[\mathbb{1}(T_1 > X_D)L_0].$$

Expression of conditional likelihood L_0 is given as (4.5) in Section

4.2.3. Then, the first term is derived as

$$\begin{aligned}
&\frac{E[\mathbb{1}(T_1 = s)\mathbb{1}(X_I < T_1 \leq X_D)L_0]}{dH_I^{(k+1)}(s)} \\
&= \frac{\mu_1\mu_2 dH_I(s)dH_I(X_I)e^{-\mu_1 H_I(s)-\mu_2 H_I(X_I)-\eta[H_D(X_D)-H_D(s)]}}{dH_I^{(k+1)}(s)} \\
&= \frac{dH_I(s)}{dH_I^{(k+1)}(s)}\mu_1\mu_2 dH_I(X_I)W(s)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}.
\end{aligned}$$

The term $E[\mathbb{1}(s \leq T_1 \leq X_D)L_0]$ can be derived as

$$\begin{aligned}
&E[\mathbb{1}(s \leq T_1 \leq X_D)L_0] \\
&= \int_{s^-}^{X_D} \mu_1\mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)+\eta H_D(x)-\mu_1 H_I(x)}dH_I(x) \\
&= \mu_1\mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}V(s^-, X_D).
\end{aligned}$$

The term $E[\mathbb{1}(T_1 > X_D)L_0]$ can be derived as

$$\begin{aligned}
E[\mathbb{1}(T_1 > X_D)L_0] &= \int_{X_D}^{\infty} \mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)}\mu_1 e^{-\mu_1 H_I(x)}dH_I(x) \\
&= \mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)-\mu_1 H_I(X_D)} \\
&= \mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}e^{\eta H_D(X_D)-\mu_1 H_I(X_D)} \\
&= \mu_2 dH_I(X_I)W(X_D)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}.
\end{aligned}$$

Then,

$$\begin{aligned}
E[U_{0,dH_I}(s)L_0] = &\frac{dH_I(s)}{dH_I^{(k+1)}(s)}\mu_1\mu_2 dH_I(X_I)W(s)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)} \\
&- \mu_1^2\mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}V(s^-, X_D) \\
&- \mu_1\mu_2 dH_I(X_I)W(X_D)e^{-\mu_2 H_I(X_I)-\eta H_D(X_D)}.
\end{aligned}$$

Thus, the unconditional score function for $dH_I(s)$

$$\begin{aligned} U_{dH_I(s)} &= \frac{E[U_{0,dH_I}(s)L_0]}{E[L_0]} \\ &= \mu_1\mu_2 \frac{\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)}W(s) - \mu_1 V(s^-, X_D) - W(X_D)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)}. \end{aligned}$$

- To calculate $U_{dH_D}(s)$, first

$$U_{0,dH_D}(s) = \begin{cases} -\eta, & T_1 = X_I, \\ -\eta + \eta \mathbb{1}(T_1 \geq s), & T_1 \in (X_I, X_D], \\ 0, & T_1 > X_I. \end{cases}$$

Then the numerator

$$\begin{aligned} E[U_{0,dH_D}(s)L_0] =& -\eta E[\mathbb{1}(T_1 = X_I)L_0] - \eta E[\mathbb{1}(X_I \leq T_1 \leq X_D)L_0] \\ &+ \eta E[\mathbb{1}(T_1 \geq s)\mathbb{1}(X_I < T_1 \leq X_D)L_0] \\ =& -\eta\mu_1 dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} W(X_I) \\ &- \eta\mu_1\mu_2 dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} V(X_I, X_D) \\ &+ \eta\mu_1\mu_2 dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} V(s^-, X_D) \\ =& -\eta\mu_1 dH_I(X_I) e^{-\mu_2 H_I(X_I) - \eta H_D(X_D)} \big[W(X_I) + \mu_2 V(X_I, s^-)\big]. \end{aligned}$$

Thus, the unconditional score function for $dH_D(s)$

$$U_{dH_D(s)} = -\eta\mu_1 \frac{W(X_I) + \mu_2 V(X_I, s^-)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)}.$$

(c) $s > X_D$, i.e. $Y_I(s) = 0$, $Y_D(s) = 0$

- To calculate $U_{dH_I}(s)$, first

$$U_{0,dH_I}(s) = \begin{cases} 0, & T_1 \leq X_D, \\ \frac{\mathbb{1}(T_1=s)}{dH_I^{(k+1)}(s)} - \mu_1 \mathbb{1}(T_1 \geq s), & T_1 > X_D. \end{cases}$$

Then, the numerator $E[U_{0,dH_I}(s)L_0]$ is

$$\frac{E[\mathbb{1}(T_1 = s)\mathbb{1}(s > X_D)L_0]}{dH_I^{(k+1)}(s)} - \mu_1 E[\mathbb{1}(T_1 \geq s)\mathbb{1}(s > X_D)L_0].$$

The term $E[\mathbb{1}(T_1 = s)\mathbb{1}(s > X_D)L_0]$ can be derived as

$$E[\mathbb{1}(T_1 = s)\mathbb{1}(s > X_D)L_0] = \mu_1\mu_2 dH_I(s)dH_I(X_I)e^{-\mu_1 H_I(s)-\mu_2 H_I(X_I)},$$

and the term $E[\mathbb{1}(T_1 \geq s)\mathbb{1}(s > X_D)L_0]$ can be derived as

$$\begin{aligned} & E[\mathbb{1}(T_1 \geq s)\mathbb{1}(s > X_D)L_0] \\ &= \mu_1\mu_2 dH_I(X_I)e^{-\mu_2 H_I(X_I)} \int_s^\infty e^{-\mu_1 H_I(x)} dH_I(x) \\ &= \mu_2 dH_I(X_I)e^{-\mu_1 H_I(s)-\mu_2 H_I(X_I)}. \end{aligned}$$

Then,

$$E[U_{0,dH_I}(s)L_0] = \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right]\mu_1\mu_2 dH_I(X_I)e^{-\mu_1 H_I(s)-\mu_2 H_I(X_I)}.$$

Thus, the unconditional score function for $dH_I(s)$

$$U_{dH_I(s)} = \mu_1\mu_2 \frac{\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} e^{\eta H_D(X_D)-\mu_1 H_I(s)}}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)}.$$

- To calculate $U_{dH_D}(s)$, since $U_{0,dH_D}(s) = 0$, $U_{dH_D}(s)$ is also zero.

Combining these results, the contribution of a subject, with intermediate event observed at X_I yet terminal event censored at X_D, to the unconditional score functions can be written as

$$\begin{aligned}
U_{dH_I}(s) &= \frac{dN_I(s)}{dH_I^{(k+1)}(s)} \\
&+ \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right]\mu_1\mu_2\frac{Y_D(s)[1-Y_I(s)]W(s) + [1-Y_D(s)]e^{\eta H_D(X_D)-\mu_1 H_I(s)}}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)} \\
-\Bigg\{Y_I(s)(\mu_1+\mu_2)&+Y_D(s)[1-Y_I(s)]\mu_1\mu_2\frac{\mu_1 V(s^-, X_D) + W(X_D) - W(s)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)}\Bigg\}, \\
U_{dH_D(s)} &= -Y_D(s)[1-Y_I(s)]\eta\mu_1\frac{W(X_I) + \mu_2 V(X_I, s^-)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + \mu_2 W(X_D)}.
\end{aligned} \tag{C.2}$$

3. Subject is censored at X_I before any event is observed (i.e. $\Delta_I = 0$, $\Delta_D = 0$)

   In this case, the denominator $E[L_0^{(k)}]$ is the observed data likelihood (4.9), so

$$E[L_0] = L_{obs,00} = e^{-(\mu_1+\mu_2)H_I(X_I)}.$$

   Then, we calculate the numerators $E[U_{0,dH_I}(s)L_0]$ and $E[U_{0,dH_D}(s)L_0]$, and finally obtain the unconditional score functions $U_{dH_I}(s)$ and $U_{dH_D}(s)$.

   Since $U_{0,dH_D(s)} = 0$ when $\Delta_I = \Delta_D = 0$, we can obtain that

$$U_{dH_D}(s) = \frac{E[U_{0,dH_D}(s)L_0]}{E[L_0]} = 0.$$

   For the calculation of $U_{dH_I}(s)$, we consider two cases: $s \le X_I$ and $s > X_I$.

   (a) $s \le X_I$, i.e. $Y_I(s) = Y_D(s) = 1$

   We first consider the conditional score function for $dH_I(s)$

$$U_{0,dH_I}(s) = -(\mu_1 + \mu_2),$$

which does not depend on the unobserved T_1.

Then, the unconditional score function for $dH_I(s)$

$$U_{dH_I}(s) = U_{0,dH_I}(s) = -(\mu_1 + \mu_2).$$

(b) $s > X_I$, i.e. $Y_I(s) = Y_D(s) = 0$

First, the conditional score function for $dH_I(s)$

$$U_{0,dH_I}(s) = \frac{\mathbb{1}(T_1 = s)}{dH_I^{(k+1)}(s)} - \mu_1 \mathbb{1}(T_1 \geq s), \quad T_1 > X_I$$

Next the numerator

$$E[U_{0,dH_I}(s)L_0] = \frac{E[\mathbb{1}(T_1 = s)L_0]}{dH_I^{(k+1)}(s)} - \mu_1 E[\mathbb{1}(T_1 \geq s)L_0].$$

The term $E[\mathbb{1}(T_1 = s)L_0]$ can be written as

$$E[\mathbb{1}(T_1 = s)L_0] = \mu_1 dH_I(s) e^{-\mu_1 H_I(s) - \mu_2 H_I(X_I)}.$$

The term $E[\mathbb{1}(T_1 \geq s)L_0]$ can be derived as

$$E[\mathbb{1}(T_1 \geq s)L_0] = \int_s^\infty \mu_1 e^{-\mu_1 H_I(x) - \mu_2 H_I(X_I)} dH_I(x) = e^{-\mu_1 H_I(s) - \mu_2 H_I(X_I)}.$$

Then, the numerator can be calculated as

$$\begin{aligned} E[U_{0,dH_I}(s)L_0] &= \frac{\mu_1 dH_I^{(k)}(s) e^{-\mu_1 H_I(s) - \mu_2 H_I(X_I)}}{dH_I^{(k+1)}(s)} - \mu_1 e^{-\mu_1 H_I(s) - \mu_2 H_I(X_I)} \\ &= \Big[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\Big] \mu_1 e^{-\mu_1 H_I(s) - \mu_2 H_I(X_I)} \end{aligned}$$

Thus, the unconditional score function for $dH_I(s)$ can be derived as

$$\begin{aligned} U_{dH_I}(s) =& \Big[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\Big]\mu_1 e^{-\mu_1 H_I(s)-\mu_2 H_I(X_I)+(\mu_1+\mu_2)H_I(X_I)} \\ =& \Big[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\Big]\mu_1 e^{\mu_1 H_I(X_I)-\mu_1 H_I(s)}. \end{aligned}$$

Combining these results, the contribution of a subject, censored at X_I before any event is observed, to the unconditional score functions can be written as

$$\begin{aligned} U_{dH_I}(s) &= -Y_I(s)(\mu_1+\mu_2) + \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right][1-Y_D(s)]\mu_1 e^{\mu_1 H_I(X_I)-\mu_1 H_I(s)}, \\ U_{dH_D}(s) &= 0. \end{aligned} \tag{C.3}$$

Combining equations (C.1), (C.2), and (C.3), we have for the unconditional score functions

$$\begin{aligned} U_{dH_I}(s) &= \frac{dN_I(s)}{dH_I^{(k+1)}(s)} - \Psi_I(s) + \left[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1\right]\theta_I(s), \\ U_{dH_D}(s) &= \frac{dN_D(s)}{dH_D^{(k+1)}(s)} - \Psi_D(s), \end{aligned}$$

where

$$\begin{aligned} \Psi_I(s) &= Y_I(s)(\mu_1+\mu_2) \\ &\quad + \Delta_I Y_D(s)[1-Y_I(s)]\mu_1\mu_2 \frac{\mu_1 V(s^-, X_D) - W(s) + (1-\Delta_D)W(X_D)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}, \\ \theta_I(s) &= (1-\Delta_I)[1-Y_D(s)]\mu_1 e^{\mu_1 H_I(X_I)-\mu_1 H_I(s)} \\ &\quad + \Delta_I\mu_1\mu_2 \frac{Y_D(s)[1-Y_I(s)]W(s) + [1-Y_D(s)](1-\Delta_D)e^{\eta H_D(X_D)-\mu_1 H_I(s)}}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}, \\ \Psi_D(s) &= \Delta_I Y_D(s)[1-Y_I(s)]\eta\mu_1 \frac{W(X_I) + \mu_2 V(X_I, s^-)}{\mu_1 W(X_I) + \mu_1\mu_2 V(X_I, X_D) + (1-\Delta_D)\mu_2 W(X_D)}. \end{aligned}$$

C.5.3 M step

Suppose we have n independent subjects with observed data $(X_{Ii}, \Delta_{Ii}, X_{Di}, \Delta_{Di})$, $i = 1, 2, ..., n$. The estimator of $dH_I^{(k+1)}(s)$ can be derived by solving $\sum_{i=1}^{n} U_{i,dH_I}(s) = 0$. Now we want to solve

$$\sum_{i=1}^{n} \left\{ \frac{dN_{Ii}(s)}{dH_I^{(k+1)}(s)} - \Psi_{Ii}(s) + \Big[\frac{dH_I^{(k)}(s)}{dH_I^{(k+1)}(s)} - 1 \Big] \theta_{Ii}(s) \right\} = 0,$$

which gives

$$dH_I^{(k+1)}(s) = \frac{\sum_i dN_{Ii}(s) + \Big[\sum_i \theta_{Ii}^{(k)}(s) \Big] dH_I^{(k)}(s)}{\sum_i [\Psi_{Ii}^{(k)}(s) + \theta_{Ii}^{(k)}(s)]}.$$

The above equation solves for $dH_I(s)$ iteratively until convergence, and the estimator at convergence $\widehat{dH}_I(s)$ is a consistent estimator of $dH_I(s)$ (Tsodikov (2003)).
Similar steps are taken to obtain the estimator of $dH_D(s)$, by solving

$$\sum_{i=1}^{n} U_{i,dH_D}(s) = \sum_{i=1}^{n} \left\{ \frac{dN_{Di}(s)}{dH_D^{(k+1)}(s)} - \Psi_{Di}(s) \right\} = 0.$$

The solution results in a Breslow-type estimator

$$dH_D^{(k+1)}(s) = \frac{\sum_i dN_{Di}(s)}{\sum_i \Psi_{Di}^{(k)}(s)},$$

which can be solved iteratively.

C.6 Asymptotic Properties

This section presents the technical details of the asymptotic properties of the NPMLE $\hat{\Omega} = (\hat{\beta}, \{d\hat{H}_I\}, \{d\hat{H}_D\})$, adapted from Hu and Tsodikov (2014), Supplementary Materials C and Rice and Tsodikov (2017), Appendix F. To establish the asymptotic properties, we assume the following conditions are satisfied (Fleming and

Harrington (2011), p289-p290):

1. The true values of baseline hazards H_I^0 and H_D^0 are strictly increasing and differentiable. The true parameter set Ω^0 is in the interior of the compact convex set $\mathcal{H}$.

2. With probability 1, $P[Y_D(t)|z] > 0$, $P[\Delta_D = 0, X_D = \tau|z] > 0$. The at risk set will not shrink to zero.

3. The Hessian matrix $\mathcal{I}_n$ evaluated at the true value $\Omega^0 = (\beta^0, dH_I^0, dH_D^0)$ is positive definite, and converges in probability to a deterministic and invertible operator $\mathcal{I}^0$.

4. The model is identifiable such that

$$(\Lambda_I, \Lambda_D) = (\Lambda_I^0, \Lambda_D^0) \quad \text{uniformly over } \Omega \quad \Rightarrow \quad \Omega = \Omega^0.$$

C.6.1 Proof of Theorem IV.1

To prove consistency: $|\hat{\Omega} - \Omega^0| \xrightarrow{p} 0$, based on Theorom 2.12 of Kosorok (2008), in addition to the regularity conditions provided above, another three conditions need to be verified:

(a) $l_n(\hat{\Omega}_n) = \sup_{\Omega \in \mathcal{H}} l_n(\Omega) - o_p(1)$

(b) Identifiability condition 2: For any sequence $\Omega_n \in \mathcal{H}$,

$$\liminf_{n\to\infty} l(\Omega_n) \geq l(\Omega^0) \quad \Rightarrow \quad \|\Omega_n - \Omega^0\| \xrightarrow{p} 0.$$

(c) Uniform convergence condition: $\sup_{\Omega \in \mathcal{H}} |l_n(\Omega) - l(\Omega)| \xrightarrow{p} 0$.

Note, $\hat{\Omega}_n$ here is the maximum likelihood estimator of Ω, which is $\hat{\Omega}$ in our previous notation.

We verify the three conditions as follows:

1. To verify condition (a), $\hat{\Omega}_n$ is the maximum estimator of Ω, thus condition (a) is satisfied.

2. To verify condition (b), by Lemma 14.3 of Kosorok (2008), we just need to prove that $l(\Omega)$ is upper semicontinuous with a unique maximum at Ω^0.

   The model is characterized by defining the hazard functions as

$$d\Lambda_I(t) = \Theta_I(t;\Omega)dH_I(t) \quad \text{and} \quad d\Lambda_D(t) = \Theta_D(t;\Omega)dH_D(t),$$

   which are both functions of Ω. Let $F_I(t)$ and $F_D(t)$ be the cumulative density functions for incidence and death, subject to censoring; and let $S_I(t)$ and $S_D(t)$ be the survival functions for observed incidence and death. $F_I^0(t)$, $F_D^0(t)$, $S_I^0(t)$ and $S_D^0(t)$ denote the corresponding true functions. Note that $dF_I(t) = S_I(t)d\Lambda_I(t)$ and $dF_D(t) = S_D(t)d\Lambda_D(t)$. The true log-likelihood can be written as

$$l(\Omega) = E\int_0^\tau \Big\{ \log d\Lambda_I(t)dF_I^0(t) - S_I^0(t)d\Lambda_I(t) + \log d\Lambda_D(t)dF_D^0(t) - S_D^0(t)d\Lambda_D(t) \Big\},$$

   where the expectation is taken over covariate $\mathbf{z}$.

   Now consider the negative Kullback-Leibler distance,

$$\begin{aligned} D &= l(\Omega) - l(\Omega^0) \\ &= E\int_0^\tau \Big\{ dF_I^0(t)[\log d\Lambda_I(t) - \log d\Lambda_I^0(t)] - S_I^0(t)[d\Lambda_I(t) - d\Lambda_I^0(t)] \\ &\qquad + dF_D^0(t)[\log d\Lambda_D(t) - \log d\Lambda_D^0(t)] - S_D^0(t)[d\Lambda_D(t) - d\Lambda_D^0(t)] \Big\}. \end{aligned}$$

Since

$$
\begin{aligned}
&dF^0(t)[\log d\Lambda(t) - \log d\Lambda^0(t)] - S^0(t)[d\Lambda(t) - d\Lambda^0(t)] \\
&= dF^0(t) \log \frac{d\Lambda(t)}{d\Lambda^0(t)} - \frac{dF^0(t)}{d\Lambda^0(t)}[d\Lambda(t) - d\Lambda^0(t)] \\
&= dF^0(t)\Big[\log \frac{d\Lambda(t)}{d\Lambda^0(t)} - \frac{d\Lambda(t)}{d\Lambda^0(t)} + 1\Big],
\end{aligned}
$$

then

$$
\begin{aligned}
D &= E \int_0^\tau \Big\{ dF_I^0(t)\big[\log \frac{d\Lambda_I(t)}{d\Lambda_I^0(t)} - \frac{d\Lambda_I(t)}{d\Lambda_I^0(t)} + 1\big] \\
&\qquad + dF_D^0(t)\big[\log \frac{d\Lambda_D(t)}{d\Lambda_D^0(t)} - \frac{d\Lambda_D(t)}{d\Lambda_D^0(t)} + 1\big] \Big\} \\
&= E \int_0^\tau \left\{ \rho\Big(\frac{d\Lambda_I(t)}{d\Lambda_I^0(t)}\Big) dF_I^0(t) + \rho\Big(\frac{d\Lambda_D(t)}{d\Lambda_D^0(t)}\Big) dF_D^0(t) \right\},
\end{aligned}
$$

where $\rho(x) = \log x - x + 1$ is a non-positive convex function, with a unique maximizer at $x = 1$. Therefore, D has a unique maximum at $\dfrac{d\Lambda_I(t)}{d\Lambda_I^0(t)} = 1$ and $\dfrac{d\Lambda_D(t)}{d\Lambda_D^0(t)} = 1$, i.e., $d\Lambda_I(t) = d\Lambda_I^0(t)$ and $d\Lambda_D(t) = d\Lambda_D^0(t)$.
Given regularity condition 4, under the identifiable model, D has a unique maximum at Ω^0. Since maximizing D is equivalent to maximizing $l(\Omega)$, we can conclude that $l(\Omega)$ has a unique maximum at Ω^0. We also know $l(\Omega)$ is upper semicontinuous, thus condition (b) holds.

3. To verify condition (c), given regularity condition 1, Ω is in the class of functions of bounded variation with integrable envelope, so $H_I(t)$ and $H_D(t)$ are bounded. Therefore, $\mathcal{H}$ is a Glivenko-Cantelli class. Then, since the functionals Λ_I, Λ_D and $l(\Omega)$ are continuous, and the envelope function is integrable, then by the preservation of Glivenko-Cantelli theorem (Van Der Vaart and Wellner (2000)), the integrand in $l(\Omega)$ is also Glivenko-Cantelli. Therefore, we apply the uniform

law of large numbers for the empirical process, such that

$$\sup_{\Omega \in \mathcal{H}} |l_n(\Omega) - l(\Omega)| \xrightarrow{p} 0.$$

C.6.2 Martingale Representation of Score Functions

In this part, we justify that the score functions for $H_I(t)$, $H_D(t)$ and β are all martingales under the true model.

Based on equations (4.12) and (4.13), we integrate the expression over $dH_I(s)$ and $dH_D(s)$, respectively, normalize them by $1/n$, and obtain the normalized score functions for $H_I(x)$ and $H_D(x)$ as

$$U_{H_I(x)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^x \left\{ \frac{dM_{Ii}(s)}{dH_I(s)} + \int_s^\tau \frac{\dot{\Theta}_{Di,dH_I}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\} dH_I(s)$$
$$U_{H_D(x)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^x \left\{ \frac{dM_{Di}(s)}{dH_D(s)} + \int_s^\tau \frac{\dot{\Theta}_{Di,dH_D}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\} dH_D(s)$$

Exchange the integrals over s and t, then we have

$$U_{H_I(x)} = \frac{1}{n}\sum_{i=1}^{n}\left\{ \int_0^\tau \mathbb{1}(t \le x) dM_{Ii}(t) + \int_0^\tau \frac{\dot{\Theta}_{Di,dH_I}(t)}{\Theta_{Di}(t)} H_I(x \wedge t) dM_{Di}(t) \right\},$$
$$U_{H_D(x)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{ \frac{\dot{\Theta}_{Di,dH_D}(t)}{\Theta_{Di}(t)} H_D(x \wedge t) + \mathbb{1}(t < x) \right\} dM_{Di}(t),$$

We first consider the score function $U_{H_I(x)}$. Similar as the proof in (Hu and Tsodikov (2014), Supplementary Materials B),

Let $\epsilon_I(t,x) = \frac{\dot{\Theta}_{D,dH_I}(t)}{\Theta_D(t)} H_I(x \wedge t)$, $\epsilon_D(t,x) = \frac{\dot{\Theta}_{D,dH_D}(t)}{\Theta_D(t)} H_D(x \wedge t) + \mathbb{1}(t < x)$, then

$$U_{H_I(x)} = \frac{1}{n} \sum_{i=1}^{n} \Big\{ \int_0^\tau \mathbb{1}(t \le x) dM_{Ii}(t) + \int_0^\tau \epsilon_{Ii}(t,x) dM_{Di}(t) \Big\}, \tag{C.4}$$

$$U_{H_D(x)} = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \epsilon_{Di}(t,x) dM_{Di}(t). \tag{C.5}$$

Consider the increment of $U_{H_I(x)}$ over x,

$$\begin{aligned} dU_{H_I}(x) = U_{H_I(x+dx)} - U_{H_I(x)} &= dM_I(x) + \int_0^\tau \epsilon_I(t, x+dx) - \epsilon_I(t,x) dM_D(t) \\ &= dM_I(x) + \int_0^\tau \frac{\partial \epsilon_I(t,x)}{\partial x} dx dM_D(t) \end{aligned}$$

Take the expectation conditional on filtration $\mathcal{F}(x^-)$,

$$E[dU_{H_I}(x)|\mathcal{F}(x^-)] = E[dM_I(x)|\mathcal{F}(x^-)] + E\Big[\int_0^\tau \frac{\partial \epsilon_I(t,x)}{\partial x} dx dM_D(t) | \mathcal{F}(x^-) \Big]$$

The first term $E[dM_I(x)|\mathcal{F}(x^-)] = 0$, based on the martingale property of $M_I(x)$. $\epsilon_I(t,x)$ depends on t when $t < x$, $\frac{\partial \epsilon_I(t,x)}{\partial x} = 0$ when $t < x$. $E[dM_D(t)|\mathcal{F}(x^-)] = 0$, if $t \ge x^-$, based on the martingale property of M_D. So the second term

$$\begin{aligned} & E\Big[\int_0^\tau \frac{\partial \epsilon_I(t,x)}{\partial x} dx dM_D(t) | \mathcal{F}(x^-) \Big] \\ = & \int_0^{x^-} E\Big[\frac{\partial \epsilon_I(t,x)}{\partial x} dx dM_D(t) | \mathcal{F}(x^-) \Big] + \int_{x^-}^\tau E\Big[\frac{\partial \epsilon_I(t,x)}{\partial x} dx dM_D(t) | \mathcal{F}(x^-) \Big] \\ = & 0 + \int_{x^-}^\tau \frac{\partial \epsilon_I(t,x)}{\partial x} dx E[dM_D(t)|\mathcal{F}(x^-)] = 0. \end{aligned}$$

Thus, $E[dU_{H_I}(x)|\mathcal{F}(x^-)] = 0$, and $U_{H_I}(x)$ is a martingale.

Similarly, we can also conclude that $U_{H_D}(x)$ is a martingale under the true model.

In terms of the score function for β, as expressed in equation (4.14), and after normalization by $\frac{1}{n}$, U_β becomes

$$U_\beta = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{ \frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t) + \frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\} \tag{C.6}$$

Since $\frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)}$ and $\frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)}$ are both predictable, so the linear transformation $\int_0^\tau \left\{ \frac{\dot{\Theta}_{Ii,\beta}(t)}{\Theta_{Ii}(t)} dM_{Ii}(t) + \frac{\dot{\Theta}_{Di,\beta}(t)}{\Theta_{Di}(t)} dM_{Di}(t) \right\}$ is a martingale. Thus, the score function with respect to β is also a martingale under the true model.

C.6.3 Proof of Theorem IV.2

We prove Theorem IV.2 in two steps. Suppose $U(\Omega) = (U_\beta, U_{H_I(t)}, U_{H_D(t)})^T$ is the set of score functions for parameter set Ω. First, we prove the weak convergence of the score functions at the true parameter $n^{1/2}U(\Omega^0)$ by Martingale Central Limit Theorem (MCLT). Then, we seek for the relationship between $n^{1/2}(\hat{\Omega} - \Omega^0)$ and $n^{1/2}U(\Omega^0)$, to obtain the weak convergence of the NPMLE $\hat{\Omega}$.

Based on the martingale representation of $U(\Omega^0)$ in Appendix C.6.2, and the fact that N_{Ii} and N_{Di}, $i = 1, 2, ..., n$, are orthogonal, it follows that $n^{1/2}U(\Omega^0)$ converges weakly to a zero-mean Gaussian process with its variance-covariance function characterized by $\sigma^2_\beta(\Omega^0)$, $\sigma^2_{H_I}(x, y; \Omega^0)$, $\sigma^2_{H_D}(x, y; \Omega^0)$, $\sigma^2_{\beta,H_I(x)}(x; \Omega^0)$, $\sigma^2_{\beta,H_D(x)}(x; \Omega^0)$ and $\sigma^2_{H_I(x),H_D(y)}(x, y; \Omega^0)$ as derived below.

The predictable variation process for score function $n^{1/2}U_\beta$ is

$$
\begin{aligned}
&< n^{1/2}U_\beta > \\
&= n\frac{1}{n^2}\sum_{i=1}^{n}\int_0^\tau \left\{ \frac{\dot{\Theta}^2_{Ii,\beta}(x)}{\Theta^2_{Ii}(x)}\Theta_{Ii}(x)Y_{Ii}(t)dH_I(t) + \frac{\dot{\Theta}^2_{Di,\beta}(t)}{\Theta^2_{Di}(t)}\Theta_{Di}(t)Y_{Di}(t)dH_D(t) \right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{ \frac{\dot{\Theta}^2_{Ii,\beta}(t)}{\Theta_{Ii}(t)}Y_{Ii}(t)dH_I(t) + \frac{\dot{\Theta}^2_{Di,\beta}(t)}{\Theta_{Di}(t)}Y_{Di}(t)dH_D(t) \right\} \\
&\xrightarrow{p} \int_0^\tau \left\{ \frac{\dot{\Theta}^2_{I,\beta}(t)}{\Theta_I(t)}P(T_I \geq t)dH_I(t) + \frac{\dot{\Theta}^2_{D,\beta}(t)}{\Theta_D(t)}P(T_D \geq t)dH_D(t) \right\}.
\end{aligned}
$$

Thus, as $n \to \infty$, $n^{1/2}U_\beta$ converges weakly to a zero-mean Gaussian process with covariance function

$$
\sigma^2_\beta(\Omega^0) = \int_0^\tau \left\{ \frac{\dot{\Theta}^2_{I,\beta}(t)}{\Theta_I(t)}P(T_I \geq t)dH_I(t) + \frac{\dot{\Theta}^2_{D,\beta}(t)}{\Theta_D(t)}P(T_D \geq t)dH_D(t) \right\}.
$$

Similarly, as $n \to \infty$,

$n^{1/2}U_{H_I}$ converges weakly to a zero-mean Gaussian process with covariance function

$$
\begin{aligned}
&\sigma^2_{H_I}(x, y; \Omega^0) \\
&= \int_0^\tau \left\{ \mathbb{1}(t \leq x)\mathbb{1}(t \leq y)P(T_I \geq t)\Theta_I(t)dH_I(t) + \epsilon_I(t, x)\epsilon_I(t, y)P(T_D \geq t)\Theta_D(t)dH_D(t) \right\},
\end{aligned}
$$

for $x, y \in [0, \tau]$;

$n^{1/2}U_{H_D}$ converges weakly to a zero-mean Gaussian process with covariance function

$$
\sigma^2_{H_D}(x, y; \Omega^0) = \int_0^\tau \epsilon_D(t, x)\epsilon_D(t, y)P(T_D \geq t)\Theta_D(t)dH_D(t);
$$

$n^{1/2}U\big(\beta, H_I(x)\big)$ converges weakly to a zero-mean Gaussian process with covariance

function

$$\begin{aligned}&\sigma^2_{\beta,H_I(x)}(x;\Omega^0)\\&=\int_0^\tau\Big\{\mathbb{1}(t\le x)\dot\Theta_{I,\beta}(t)P(T_I\ge t)dH_I(t)+\epsilon_I(t,x)\dot\Theta_{D,\beta}(t)P(T_D\ge t)dH_D(t)\Big\};\end{aligned}$$

$n^{1/2}U\big(\beta,H_D(x)\big)$ converges weakly to a zero-mean Gaussian process with covariance function

$$\sigma^2_{\beta,H_D(x)}(x;\Omega^0)=\int_0^\tau\epsilon_D(t,x)\dot\Theta_{D,\beta}(t)P(T_D\ge t)dH_D(t);$$

$n^{1/2}U\big(H_I(x),H_D(y)\big)$ converges weakly to a zero-mean Gaussian process with covariance function

$$\sigma^2_{H_I(x),H_D(y)}(x,y;\Omega^0)=\int_0^\tau\epsilon_I(t,x)\epsilon_D(t,y)P(T_D\ge t)\Theta_D(t)dH_D(t).$$

Let the normalized log-likelihood $\dfrac{1}{n}\sum\limits_{i=1}^n l_i$ converges in probability to l_∞, and $U_\infty=\Big(\dfrac{\partial l_\infty}{\partial\beta},\dfrac{\partial l_\infty}{\partial dH(t)}\Big)^T$, where $dH(t)=(dH_I(t),dH_D(t))$. Define a linear information operator $\mathcal{I}_\infty$ as

$$\mathcal{I}_\infty(t,s)=-\frac{\partial U_\infty}{\partial\Omega}\Big|_{\Omega=\Omega^0}=-\begin{bmatrix}\frac{\partial^2 l_\infty}{\partial\beta\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)}\\ \frac{\partial^2 l_\infty}{\partial dH(t)\partial\beta^T} & \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)}\end{bmatrix}_{\Omega=\Omega^0},$$

and the operator $\mathcal{I}_\infty$ acts on an arbitrary vector-function element $\Omega_s=\big(\beta,dH(s)\big)^T$ as follows

$$\mathcal{I}_\infty(t,s)\Omega_s=-\begin{bmatrix}\frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)}\big|_{\Omega=\Omega^0}\beta+\int_0^\tau\frac{\partial^2 l_\infty}{\partial\beta\partial dH(s)}\big|_{\Omega=\Omega^0}dH(s)\\ \frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)}\big|_{\Omega=\Omega^0}\beta+\int_0^\tau\frac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)}\big|_{\Omega=\Omega^0}(s)\end{bmatrix}.$$

With Taylor expansion, expand $U(\hat{\Omega})$ at the true parameter Ω^0, we have

$$\begin{aligned} & U(\hat{\Omega}) = 0 = U(\Omega^0) - \mathcal{I}_\infty(t,s)(\hat{\Omega} - \Omega^0) + o_p(1) \\ \Rightarrow \quad & \mathcal{I}_\infty(t,s) n^{1/2}(\hat{\Omega} - \Omega^0) = n^{1/2} U(\Omega^0) + o_p(1). \end{aligned} \tag{C.7}$$

Assume that the Fredholm operator expressed by the kernel $\mathcal{I}_\infty$ of the Fredholm integral equations (C.7) of the first kind is square integrable, and the equation $\mathcal{I}_\infty \Omega = 0$ has only the trivial solution $\Omega = 0$. Then based on Theorem 3.3.1 of Vaart and Wellner (1996), equations (C.7) has the unique solution, and there exists the inverse information operator $\mathcal{I}_\infty^{-1}(t,s)$ such that

$$n^{1/2}(\hat{\Omega} - \Omega^0) = \mathcal{I}_\infty^{-1}(t,s) n^{1/2} U(\Omega^0) + o_p(1)$$

Take differentiation of the equation $E[U(\Omega^0)] = 0$ with respect to Ω at the true parameter Ω^0, we have the equivalence between $\mathcal{I}_\infty$ represented by the log-likelihood second derivative, and

$$\mathcal{I}_\infty(t,s) = \begin{bmatrix} \frac{\partial l_\infty}{\partial \beta} \frac{\partial l_\infty}{\partial \beta^T} & \frac{\partial l_\infty}{\partial \beta} \frac{\partial l_\infty}{\partial dH(s)} \\ \frac{\partial l_\infty}{\partial dH(t)} \frac{\partial l_\infty}{\partial \beta^T} & \frac{\partial l_\infty}{\partial dH(t)} \frac{\partial l_\infty}{\partial dH(s)} \end{bmatrix}_{\Omega = \Omega^0},$$

which represents the variance of the score process $n^{1/2} U(\Omega^0)$. In addition, Andersen et al. (2012) showed that for a differentiable functional $F(\Omega)$, by functional delta method, $n^{1/2}\{F(\hat{\Omega}) - F(\Omega)\}$ converges weakly to a zero-mean Gaussian process with variance-covariance function $\dot{F}(\Omega)^T I_\infty^{-1} \dot{F}(\Omega)$, where $\dot{F}(\Omega) = \frac{\partial F}{\partial \Omega}$. Applying it to the linear functional defined in Theorem IV.2, and replacing I_∞ by its consistent estimator $\mathcal{I}_n$, we can have the stated results.

C.6.4 Proof of Theorem IV.3

As we defined in the proof of Theorem IV.2,

$$\mathcal{I}_\infty(t,s) = -\begin{bmatrix} \dfrac{\partial^2 l_\infty}{\partial\beta\partial\beta^T} & \dfrac{\partial^2 l_\infty}{\partial\beta\partial dH(s)} \\ \dfrac{\partial^2 l_\infty}{\partial dH(t)\partial\beta^T} & \dfrac{\partial^2 l_\infty}{\partial dH(t)\partial dH(s)} \end{bmatrix}_{\Omega=\Omega^0} = \begin{bmatrix} \mathcal{I}_{\beta\beta} & \mathcal{I}_{\beta H} \\ \mathcal{I}_{H\beta} & \mathcal{I}_{HH} \end{bmatrix},$$

is the asymptotic covariance matrix operator of the score for the full model. Apply the formula of block matrix inverse to the matrix operator, then we have the inverse of $\mathcal{I}_\infty$ as

$$\mathcal{I}_\infty^{-1} = \begin{bmatrix} Q^{-1} & -Q^{-1}\mathcal{I}_{H\beta}\mathcal{I}_{HH}^{-1} \\ -\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}Q^{-1} & \mathcal{I}_{HH}^{-1} + \mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}Q^{-1}\mathcal{I}_{\beta H}\mathcal{I}_{HH}^{-1} \end{bmatrix},$$

where $Q = \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta H}\mathcal{I}_{HH}^{-1}\mathcal{I}_{H\beta}$.

In this part, we need to justify $I_{\beta\beta}^{pr} = \left(-\frac{\partial^2 l_{pr,n}}{\partial\beta\partial\beta^T}\Big|_{\beta=\hat\beta} \right)^{-1} \xrightarrow{p} \text{var}\left[\sqrt{n}(\hat\beta - \beta^0)\right]$, where $l_{pr,n} = n^{-1}\sum_{i=1}^n l_{pr,i}(\beta) = n^{-1}\sum_{i=1}^n l_i\big(\beta, \{d\hat H(\beta)\}\big)$. We prove it in two steps: First, we show that Q^{-1},the $\beta\beta$ submatrix of $\mathcal{I}_\infty^{-1}$, is the limiting covariance matrix of $\hat\beta$, i.e., var $\left[\sqrt{n}(\hat\beta - \beta^0)\right] = Q^{-1}$. Then, we prove that $I_{\beta\beta}^{pr\,-1} \xrightarrow{p} Q^{-1}$.

First, the proof of var $\left[\sqrt{n}(\hat\beta - \beta^0)\right] = Q^{-1}$ is as follows:

Based on Appendix C.4, we have

$$U_\Omega(\Omega^0) = \frac{\partial l_n}{\partial\Omega}\Big|_{\Omega=\Omega^0} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{\partial \log d\Lambda_i(t)}{\partial\Omega} dM_i(t),$$

which is a martingale under the true model.

The predictable variation process for $\sqrt{n}U_\Omega(\Omega^0)$ is

$$< \sqrt{n}U_\Omega(\Omega^0) >= \widehat{\text{cov}}[\sqrt{n}U_\Omega(\Omega^0)] = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{\partial \log d\Lambda_i(t)}{\partial\Omega}\frac{\partial \log d\Lambda_i(t)}{\partial\Omega^T} Y_i(t) d\Lambda_i(t). \tag{C.8}$$

The negative Hessian of the profile log-likelihood can be derived as

$$
\begin{aligned}
I^{pr}_{\beta\beta} =& - \frac{d^2 l_{pr}}{d\beta d\beta^T}\Big|_{dH=d\hat{H}_\beta} \\
=& - \frac{\partial}{\partial dH(t)}\left[\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\right]\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial}{\partial \beta^T}\left[\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\right] \\
& - \frac{\partial}{\partial dH(t)}\frac{\partial l}{\partial \beta}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big|_{d\hat{H}_\beta} \\
=& - \frac{\partial^2 l}{\partial dH(s)\partial dH(t)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial \beta}\frac{\partial d\hat{H}_\beta(t)}{\partial \beta^T} - \frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial^2 d\hat{H}_\beta(s)}{\partial\beta\partial dH(t)}\frac{\partial d\hat{H}_\beta(t)}{\partial\beta^T} \\
& - \frac{\partial^2 l}{\partial dH(s)\partial\beta^T}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial\beta} - \frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial^2 d\hat{H}_\beta(s)}{\partial\beta\partial\beta^T} - \frac{\partial^2 l}{\partial\beta\partial dH(s)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(s)}{\partial\beta^T} \\
& - \frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big|_{d\hat{H}_\beta} \\
=& J_{\beta H}\hat{I}_{HH}J_{H\beta} - 0 + J_{\beta H}\hat{I}_{H\beta} - 0 + \hat{I}_{\beta H}J_{H\beta} + \hat{I}_{\beta\beta} \\
=& J_{\beta H}\hat{I}_{HH}J_{H\beta} + J_{\beta H}\hat{I}_{H\beta} + \hat{I}_{\beta H}J_{H\beta} + \hat{I}_{\beta\beta}.
\end{aligned}
$$

To express $I^{pr}_{\beta\beta}$ without the Jacobians, we make use of the fact that $\dfrac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta} = 0$, such that

$$
\begin{aligned}
& \frac{d}{d\beta}\left(\frac{\partial l}{\partial dH(s)}\Big|_{d\hat{H}_\beta}\right) = 0 \\
=& \frac{\partial^2 l}{\partial dH(s)\partial\beta}\Big|_{d\hat{H}_\beta} + \frac{\partial^2 l}{\partial dH(s)\partial dH(t)}\Big|_{d\hat{H}_\beta}\frac{\partial d\hat{H}_\beta(t)}{\partial\beta} = -\hat{I}_{H\beta} - \hat{I}_{HH}J_{H\beta} \\
\Longrightarrow\quad & J_{H\beta} = -\hat{I}^{-1}_{HH}\hat{I}^{-1}_{H\beta}
\end{aligned}
$$

Replace the Jacobians in $I^{pr}_{\beta\beta}$ with the expression above, we have

$$
I^{pr}_{\beta\beta} = \hat{I}_{\beta\beta} - \hat{I}_{\beta H}\hat{I}^{-1}_{HH}\hat{I}_{H\beta} \quad \xrightarrow{p} \quad \mathcal{I}_{\beta\beta} - \mathcal{I}_{\beta H}\mathcal{I}^{-1}_{HH}\mathcal{I}_{H\beta} = Q.
$$

Thus,

$$
I^{pr\,-1}_{\beta\beta} \xrightarrow{p} Q^{-1} = \mathrm{var}[\sqrt{n}(\hat{\beta} - \beta^0)].
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

American Cancer Society (2020). Prostate Cancer Statistics. `https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html`.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Andersen, P. K., Hansen, L. S., and Keiding, N. (1991). Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process. *Scandinavian Journal of Statistics*, pages 153–167.

Andriole, G. L., Crawford, E. D., Grubb III, R. L., Buys, S. S., Chia, D., Church, T. R., Fouad, M. N., Gelmann, E. P., Kvale, P. A., Reding, D. J., et al. (2009). Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine*, 360(13):1310–1319.

Andriole, G. L., Crawford, E. D., Grubb III, R. L., Buys, S. S., Chia, D., Church, T. R., Fouad, M. N., Isaacs, C., Kvale, P. A., Reding, D. J., et al. (2012). Prostate cancer screening in the randomized prostate, lung, colorectal, and ovarian cancer screening trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute*, 104(2):125–132.

Baker, S. G. (1994). The multinomial-poisson transformation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(4):495–504.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Chen, Y.-H. (2009). Weighted breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika*, 96(3):591–600.

Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2):235–251.

Chen, Y.-H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime data analysis*, 18(1):36–57.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Cook, R. J. and Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

Dejardin, D., Lesaffre, E., and Verbeke, G. (2010). Joint modeling of progression-free survival and death in advanced cancer clinical trials. *Statistics in Medicine*, 29(16):1724–1734.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4):907–919.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.

Force, U. P. S. T. et al. (2008). Screening for prostate cancer: Us preventive services task force recommendation statement. *Annals of internal medicine*, 149(3):185–191.

Gates, T. J. (2001). Screening for cancer: evaluating the evidence. *American family physician*, 63(3):513.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Hougaard, P. (2012). *Analysis of multivariate survival data*. Springer Science & Business Media.

Hsieh, J.-J., Wang, W., and Adam Ding, A. (2008). Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):3–20.

Hu, C. and Tsodikov, A. (2013). Semiparametric regression analysis for time-to-event marked endpoints in cancer studies. *Biostatistics*, 15(3):513–525.

Hu, C. and Tsodikov, A. (2014). Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime data analysis*, 20(4):563–583.

Huang, X. and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63(2):389–397.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Kong, S., Nan, B., Kalbfleisch, J. D., Saran, R., and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association*, 113(521):357–368.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Lang, J. B. (1996). On the comparison of multinomial and poisson log-linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):253–266.

Lin, D., Sun, W., and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika*, 86(1):59–70.

Lin, K., Lipsitz, R., Miller, T., and Janakiraman, S. (2008). Benefits and harms of prostate-specific antigen screening for prostate cancer: an evidence update for the us preventive services task force. *Annals of internal medicine*, 149(3):192–199.

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

Oakes, D. (1994). Multivariate survival distributions. *Journaltitle of Nonparametric Statistics*, 3(3-4):343–354.

Pinsky, P. F., Black, A., Parnes, H. L., Grubb, R., Crawford, E. D., Miller, A., Reding, D., and Andriole, G. (2012). Prostate cancer specific survival in the prostate, lung, colorectal, and ovarian (plco) cancer screening trial. *Cancer epidemiology*, 36(6):e401–e406.

PLCO Trial (2010). The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. `https://cdas.cancer.gov/plco/`.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21.

Rice, J. D. and Tsodikov, A. (2017). Semiparametric time-to-event modeling in the presence of a latent progression event. *Biometrics*, 73(2):463–472.

Schaubel, D. E. and Cai, J. (2004). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika*, 91(2):291–303.

SEER (2020). The surveillance, epidemiology, and end results (seer) program. `https://seer.cancer.gov/statfacts/html/prost.html`.

Shoag, J. E., Mittal, S., and Hu, J. C. (2016). Reevaluating psa testing rates in the plco trial. *The New England journal of medicine*, 374(18):1795–1796.

Shu, X. and Schaubel, D. E. (2016). Semiparametric methods to contrast gap time survival functions: Application to repeat kidney transplantation. *Biometrics*, 72(2):525–534.

Stanford, J. L., Stephenson, R. A., Coyle, L. M., Cerhan, J., Correa, R., Eley, J., Gilliland, F., Hankey, B., Kolonel, L., Kosary, C., et al. (1999). Prostate cancer trends 1973–1995, seer program, national cancer institute. *NIH pub*, (99-4543):60.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):759–774.

Tsodikov, A. and Chefo, S. (2008). Generalized self-consistency: Multinomial logit model and poisson likelihood. *Journal of statistical planning and inference*, 138(8):2380–2397.

Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer.

Van Der Vaart, A. and Wellner, J. A. (2000). Preservation theorems for glivenko-cantelli and uniform glivenko-cantelli classes. In *High dimensional probability II*, pages 115–133. Springer.

Vickers, A. J. (2017). Prostate cancer screening: time to question how to optimize the ratio of benefits and harms. *Annals of internal medicine*, 167(7):509–510.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika*, 56(3):601–614.

Zeng, D. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):507–564.

Zeng, D. and Lin, D. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica*, 20(2):871.